ISSN digital 2236-1766

Direito Público

Revista Oficial do Programa de Pós-Graduação *Stricto Sensu* em Direito – Mestrado e Doutorado Acadêmico – do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

Ano 18 – nº 100 – Out-Dez 2021

Indexada por

Diretório de Revistas Brasileiras em SEER, Diadorim – Diretório de Política de Acesso Aberto das Revistas Científicas Brasileiras, Portal de Periódicos da CAPES, Latindex, Index Copernicus Internacional, Directory of Open Access Journal – DOAJ

Uma publicação do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa – IDP.

Revisão e Diagramação: Dois Pontos Editoração

Artigos para possível publicação devem ser encaminhados exclusivamente por meio do Portal de Periódicos do IDP (www.direitopublico.idp.edu.br), com o prévio cadastramento do Autor.

Dados Internacionais de Catalogação na Publicação (CIP)

D598 Direito Público. (recurso eletrônico) / Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, v. 1, n. 1 (jul./set. 2003)-

Brasília: Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, 2005-.
v. 18, n. 100, out./dez. 2021-
Trimestral
ISSN: 2236-1766 (versão digital)

1. Direito público – periódicos. I. Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa.

CDU 342
CDU 341

(Bibliotecária responsável: Débora Souza – CRB 1/3249)

Solicita-se permuta.
Pídese canje.
On demande l'échange.
Si richiede lo scambio.
We ask for exchange.
Wir bitten um austausch.

Permuta com as Instituições:
Escola Nacional de Administração Pública. Biblioteca Graciliano Ramos.
Escola Superior da Magistratura. Ajuris. Biblioteca.
Pontifícia Universidade Católica de Minas Gerais. Contagem. Biblioteca.
Senado Federal. Biblioteca Acadêmico Luiz Viana Filho.
Universidade de Brasília. Biblioteca Central.
Universidade de Lisboa. Biblioteca.
Universidade de Santa Cruz do Sul. Biblioteca Central.
Universidade Federal de Santa Catarina. Biblioteca Universitária.
Universidade do Vale do Itajaí. Biblioteca Central Comunitária.
Universidade do Vale do Rio dos Sinos. Biblioteca.
Universidade Federal do Paraná. Biblioteca Central.
Universidade Federal do Rio Grande do Sul. Biblioteca.

IDP – Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa

www.idp.edu.br

SGAS 607 – Módulo 49 – Av. L2 Sul – Asa Sul
70766-090 – Brasília – DF
Fone/Fax: (61) 3535-6590

E-mail: direitopublico@idp.edu.br

TOTA MACHINA

Uma instigante interseção entre arte e inteligência artificial na arte de Katia Wille

Em "Das tripas coração", individual que a artista visual Katia Wille apresentou de Março a Junho no Museu da República do Rio de Janeiro, as obras de arte vêm o espectador antes mesmo do espectador ver a obra .

A ideia é estabelecer uma simbiose entre obras de arte e o público . Para que isso fosse possível, a artista desenvolveu em parceria inédita com a Microsoft um conceito de máquinas cognitivas integradas ao ambiente, que está sendo mostrado primeira vez em uma exposição de Arte no Brasil - alem da exposição no Museu da República, uma das obras acaba de ser mostrada na Casa FIRJAN durante a exposição Data Corpus e foi exposta em Junho no evento de tecnologia .Futuro|Rio. A partir de outubro a mostra segue para São Paulo.

Utilizando a capacidade de inteligência artificial na nuvem, as obras reagem a presença de pessoas, refletem sentimentos e interagem através de movimentos diante de estímulos visuais, faciais e sonoros. O projeto usa robótica e IA para análise de ambiente, sentimentos e voz, conectando o público com as obras e proporcionando uma experiência única a cada espectador por meio da tecnologia. Cada uma das obras traz uma experiência diferente as pessoas, abrindo assim o debate sobre a crescente interseção entre arte e tecnologia.


Katia Wille
DAS
TRIPAS
CORAÇÃO
Uma experiência unindo arte e IA
Apoio:
Microsoft

"Katia Wille, em ***Das Tripas Coração*** pretende estabelecer um diálogo entre o espectador e a obra de arte, evidenciando a fragilidade das relações humanas e a busca incessante pelo olhar do outro. Partindo de inquietações e questões levantadas ao longo de sua carreira a artista procura entender o movimento de procura do interno no externo, a dinâmica complexa da relação humana com o mundo. Corpos em movimento, entrelaçados, curvados ou muito esticados são constantes em sua obra e estabelecem uma relação com a delicada questão da procura pelo equilíbrio. Misturados os corpos perdem a sua identidade para formar uma massa homogênea, semelhante às tripas e partes internas do ser humano. Os materiais escolhidos por Katia procuram expressar conceitos como transformação, elasticidade, porosidade e fragmentação através de instalações e pinturas em ecolatex e tecido. O projeto, com apoio tecnológico, se propõe a criar obras que sejam em sua essência, reflexivas. Que permitam, através do material usado, que o espectador se veja refletido na obra, e entenda que a matéria é frágil. Corpo, cérebro e pele, as três camadas ou pilares da obra dessa artista que procura ir além de conceitos estabelecidos nessa relação tão delicada que a arte possibilita."

Iabel Sanson Portella

Sobre Katia Wille

Nascida no Rio de Janeiro, Katia Wille é formada em artes e design de moda pela Universidade de Amsterdam, na Holanda, e passou os últimos 10 anos morando e trabalhando entre a Europa, a Ásia e o Brasil. As questões do feminino, do corpo em busca de sua essência e transformações, sempre povoaram as obras da artista, que pensa movimento e cor integrados ao todo.

Exposições individuais: "Das tripas coração"-Galeria do Lago /Museu da República/RJ; "Mas Afinal: Quem tem medo de tamanha liberdade?" - Galeria VillaNova/SP; "Fluxofloração" – Centro Cultural da Justiça Federal/RJ; "Maria dos olhos de piscina" – H.Contemporaneo/RJ; "O Tudo Do Todo" – Z42 Arte Contemporanea/RJ ; "E daí? Eu adoro voar" – Tramas Arte Contemporânea/RJ; "CompulsArt" – Casa Ipanema/RJ; "As Nadadoras – Livre Galeria/RJ

Exposições coletivas: "Um dia de sol" – Galeria Sal/RJ; "Arte brasileira na contemporaneidade" – InnGaleria/SP; "Para Todos" – Galeria Carpintaria/Fortes D'Aloia & Gabriel/RJ"; "Olhar Feminino" – Galeria André/SP; "Somos todos Clarisse" – Museu da República/Galeria do Lago/RJ; "A Máquina do Mundo – residência artística" - Z42/RJ; "Weel Chair Fest/ Rio Olympic Games" - Boulevard Olímpico/RJ

Publicações: "Arte brasileira na contemporaneidade" - Volume: III, Ornitorrinco - São Paulo, Brasil. Agosto 2018 Autora: Carmen Pousada

Prêmios e programas de residência artística: "European Union Visual Arts and Design Awards" - Tokyo (Japão), Março 2010;
Z42 - Rio de Janeiro, Brasil

Inteligência Artificial, Ética e Epistemologia: o Encontro da Tecnologia com as Ciências Sociais

MIRIAM WIMMER

Doutora em Políticas de Comunicação e Cultura pela Faculdade de Comunicação da Universidade de Brasília (UnB). Mestre em Direito Público e Graduada em Direito pela Universidade do Estado do Rio de Janeiro (UERJ). Ex-Bolsista do Programa Internacional da Universidade de Waseda, com Distinção Acadêmica. Certificada como Especialista em Proteção de Dados Pessoais (Europa) pela International Association of Privacy Professionals (CIPP/E). Professora do Corpo Permanente do Mestrado Profissional em Direito do IDP-Brasília e Professora Convidada em diversas instituições de ensino de nível superior. Servidora Pública desde 2007, integrante da carreira de Especialista em Regulação de Serviços Públicos de Telecomunicações da Anatel. Ocupou diferentes cargos de direção no Ministério das Comunicações – MC e no Ministério de Ciência, Tecnologia, Inovações e Comunicações – MCTIC, onde coordenou a elaboração da Estratégia Brasileira para a Transformação Digital e atuou na propositura da Estratégia Brasileira de IA. É, atualmente, Diretora da Autoridade Nacional de Proteção de Dados – ANPD.

DANILO DONEDA

Doutor em Direito Civil pela Universidade do Estado do Rio de Janeiro (UERJ). Professor no Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP). Representante da Câmara dos Deputados no Conselho Nacional de Proteção de Dados e Privacidade. Diretor do CEDIS/IDP (Centro de Estudos de Internet e Sociedade). Membro do Conselho Diretor da IAPP (International Association of Privacy Professionals). Foi Coordenador-Geral na Secretaria Nacional do Consumidor do Ministério da Justiça, onde coordenou a redação do Anteprojeto de Lei de Proteção de Dados, a base da Lei Geral de Proteção de Dados. Membro da Comissão de Juristas da Câmara dos Deputados para redação de Projeto de Lei sobre Proteção de Dados nos setores de segurança pública e investigação criminal. Membro do Grupo de Trabalho sobre Proteção de Dados e Informações Judiciais do Conselho Nacional de Justiça. Membro dos Conselhos Consultivos do Projeto Global Pulse (ONU) e do Projeto Criança e Consumo (Instituto Alana). Foi Pesquisador Visitante na Autoridade Garante para a Proteção de Dados em Roma (Roma, Itália), na Università degli Studi di Camerino (Camerino, Itália) e no Instituto Max Planck para Direito Privado Comparado e Internacional (Hamburgo, Alemanha). Parte do seu trabalho está disponível em: www.doneda.net.

Iniciar uma discussão sobre tecnologia com a afirmação de que os tempos em que vivemos são singulares é, decerto, uma fórmula bastante banalizada. Apesar do lugar-comum, a constatação não deixa de ser verdadeira: a evolução tecnológica provoca mudanças céleres e intensas em nossa sociedade, colocando em xeque estruturas institucionais, regras jurídicas e consensos sociais anteriormente válidos. Não por acaso, tais fenômenos costumam ser descritos como integrantes de uma "revolução", de uma "disrupção" ou de uma "quebra de paradigmas".

Esse processo de transformação está indubitavelmente relacionado ao acelerado desenvolvimento da inteligência artificial, tecnologia que há poucas décadas era compreendida apenas por uma seleta casta de cientistas e que agora vem migrando dos filmes de ficção científica em direção às prateleiras de supermercado. Nunca se falou tanto sobre IA, seja nas universidades, seja nas empresas, seja nas mesas de bar. E não é para menos: são significativos os avanços conquistados nos últimos anos, particularmente nos campos de *machine learning* e *deep learning*, como também inquestionável a presença cada vez mais pervasiva da IA no mundo dos negócios, no governo e no dia a dia das pessoas comuns.

Os principais vetores para essa evolução podem ser identificados. Ainda que alguns dos paradigmas e conceitos mais relevantes relacionados à IA existam há décadas, um salto mais recente veio tanto do incremento na capacidade de processamento de computadores quanto da maior disponibilidade daquilo que acaba sendo a matéria-prima de algumas das aplicações mais impactantes e promissoras dessa tecnologia, que são os dados pessoais necessários para o "treinamento" de diversos sistemas.

Ao mesmo tempo que o desenvolvimento tecnológico permite entrever importantes ganhos econômicos e benefícios sociais nos mais diversos campos da atividade humana, novas e urgentes questões se apresentam. Embora a inteligência artificial tenha inicialmente se desenvolvido, sobretudo, a partir de disciplinas das ciências exatas, as inovadoras possibilidades que se descortinam empurram o debate em direção a campos ainda pouco compreendidos, com implicações sensíveis e significativas para os seres humanos.

O cenário atual, com a efetiva e progressiva inserção de sistemas de IA em nosso cotidiano, inaugura uma nova etapa na qual as análises sobre o tema não podem mais considerar meramente perspectivas futuras de eficácia – visto que as implementações da tecnologia são concretas –, porém devem também levar em conta elementos que, antes, seriam considerados como laterais ou secundários, como os impactos da utilização de IA para os titulares de dados pessoais afetados, potenciais efeitos quanto à discriminação ou ao desequilíbrio de poder, entre outros.

Nesse contexto, as intensas discussões travadas quanto ao tema em arenas internacionais e no campo legislativo revelam as inquietações que permeiam o debate quanto a critérios regulatórios e éticos para o desenvolvimento e uso de IA, assim como a necessidade de melhor compreensão

das condições que informam as capacidades cognitivas de agentes não humanos em comparação com as formas humanas de aquisição do conhecimento.

A partir de tal contexto, o presente Dossiê Temático busca contribuir para a discussão sobre IA, ética e epistemologia, reunindo, em um único volume, artigos nacionais e internacionais que abordam diferentes dimensões dessa problemática. Buscou-se reunir um conjunto diverso de textos capazes de propiciar reflexões teóricas e práticas que possam ajudar a informar o debate público sobre o tema. O elevado número de submissões recebidas para o Dossiê e a notável qualidade das contribuições refletem a grande atenção que o tema tem recebido da comunidade científica. Tornaram, também, extremamente desafiador o trabalho dos organizadores de selecionar artigos representativos das principais questões em debate!

O Dossiê conta com quatro artigos de autores internacionais. O primeiro, intitulado "*Electronic democracy and digital justice: driving principles for AI regulation in the prism of the Human Rights*", de autoria do Professor Alessandro Mantelero, debruça-se sobre dois campos em que os impactos da IA são significativos: democracia eletrônica e justiça digital. Embora ambos os temas sejam regidos por instrumentos jurídicos tanto vinculantes como não vinculantes, os princípios que os orientam foram desenvolvidos em uma era pré-IA, o que aponta para a necessidade de que sejam reavaliados. O trabalho de Mantelero é relevante, dentre outros aspectos, por indicar a necessidade de uma integração mais equilibrada entre lei e ética na regulação de IA, tendo por pilar central os direitos humanos. Para o autor, uma contribuição efetiva para o debate sobre direitos humanos deve partir de uma adequada contextualização dos direitos e das liberdades fundamentais face à inteligência artificial.

Já o artigo "*Ethical dimensions of the GDPR, AI regulation, and beyond*", de Hielke Hijmans e Charles Raab, traz contribuição atual e significativa ao analisar, sob a ótica do regulamento europeu de proteção de dados, as novas questões éticas relacionadas à privacidade e à proteção de dados associadas aos desenvolvimentos sociotécnicos que têm por base o uso extensivo e intensivo de dados. O artigo analisa, ainda, a proposta europeia de 2021 de regulamentação da IA, apresentando, ao fim, considerações acerca do papel que diferentes atores, como organizações privadas, comitês de ética e Autoridades de Proteção de Dados, podem desempenhar na tomada de decisões sobre ética em IA.

No artigo "Vigilância e relações de poder: o uso de tecnologias de reconhecimento facial e identificação biométrica a distância em espaço público e impactos na vida pública", Ana Catarina Fontes e Christoph Lütge introduzem os conceitos de espaço público e vida pública para refletir sobre os impactos da implementação de tecnologias de vigilância intrusivas, como tecnologias de reconhecimento facial e identificação biométrica, muitas vezes introduzidas sob a justificativa de garantia da segurança e da ordem pública. As considerações dos autores tocam em questões éticas relacionadas com valores culturais e direitos adquiridos e problematizam tais tecnologias com base nos princípios de transparência, autonomia, proporcionalidade e equidade.

Fechando a área internacional do Dossiê, o artigo "*Internet and the political campaigns. The perspective of the Italian Legislation*", de Ornella Spataro, aborda o tema da influência das mídias sociais sobre a esfera pública e sua compatibilidade com a teoria democrática liberal, principalmente por modificar diversos vetores do acesso dos atores políticos aos meios de comunicação que hoje assumem maior relevância. A partir daí, examina-se a reação do ordenamento italiano a essa situação, considerando, entre outras, a aplicação da legislação eleitoral, da legislação de proteção de dados e do regulador (no caso, a AGCom, Autoridade para as garantias nas comunicações), para identificar, ao cabo, a necessidade de maiores intervenções regulatórias.

Adentrando o terreno da produção científica nacional, o Dossiê apresenta dez artigos que abordam diferentes aspectos da discussão ética e epistemológica sobre a IA.

A partir da constatação de que as filosofias tradicionais e as ciências jurídicas clássicas não se prepararam para enfrentar os dilemas da contemporaneidade, o artigo "Prolegômenos a uma filosofia algorítmica futura que possa apresentar-se como fundamento para um cyberdireito", de autoria de Mariah Brochado, lança uma provocação sobre a possibilidade de se pensar em uma filosofia algorítmica e em um cyberdireito, inspirados pelos desafios decorrentes do advento da revolução tecnológica e que passaram a ser recorrentes no cotidiano dos humanos, como o *big data*, a inteligência artificial, a rotina das redes sociais, as descobertas no campo da autonomia de máquinas e os desconfortos éticos surgidos destas e de outras experiências novas e insólitas, caracterizadas por sua radical diferença com relação ao mundo analógico.

Já o artigo "*What is it like to be an artificial intelligence? Nagel's view from nowhere and artificial intelligence as subject of Law*", de autoria de Sthéfano Bruno Santos Divino e Rodrigo Almeida Magalhães, toma como ponto de partida os aspectos epistemológicos e semânticos desenvolvidos por Thomas Nagel para interrogar: como é ser uma inteligência artificial? A partir de tal análise, os autores buscam avaliar se um ente inteligente artificialmente pode, ou não, ser considerado sujeito de direito, concluindo que, ainda que no estágio científico atual um ente inteligente artificialmente (ainda) não possa ser considerado sujeito de direito, sob pena de caracterização do instrumentalismo, a conquista de tal condição é, em tese, possível, superando-se a visão redutiva pautada no fisicalismo.

No artigo intitulado "Argumentação jurídica e aprendizado profundo", os autores Orlando Luiz Zanon Junior e Guilherme Kirtschig apresentam, a partir das concepções vinculadas a diferentes paradigmas da ciência do Direito, um panorama da argumentação jurídica e sua importância na aplicação do Direito, para, em seguida, investigar a possibilidade de sua automatização, mediante uso da tecnologia de aprendizado profundo. O artigo apresenta interessante problematização do efetivo desempenho da argumentação jurídica por parte de sistemas de IA e sopesa suas exigências com as limitações inerentes ao aprendizado profundo, argumentando quanto à possibilidade de compatibilização entre ambos.

Conectando-se com o debate precedente sobre as formas pelas quais sistemas de IA chegam a decisões, os dois artigos seguintes abordam o tema da discriminação, que é hoje um dos aspectos mais debatidos da inteligência artificial. O artigo "Todos são iguais perante o algoritmo? Uma resposta cultural do direito à discriminação algorítmica", de autoria de Alan Duarte e do Ramon de Vasconcelos Negócio, explora a forma pela qual algoritmos de aprendizado de máquina que operam em rede podem apresentar comportamentos discriminatórios, de modo a evidenciar uma semântica da discriminação que perpassa instituições e estruturas sociais, incorporando-se aos aspectos técnicos da programação e ensejando ameaças a direitos fundamentais. O artigo adota a noção de igualdade como não discriminação e parte da compreensão do Direito como uma semântica social para sugerir mecanismos de construção de um modelo de observação, identificação e busca de soluções voltadas ao combate da discriminação por algoritmos.

Já o artigo "Discriminação algorítmica e inclusão em sistemas de inteligência artificial – Uma reflexão sob a ótica dos direitos da criança no ambiente digital", de Isabella Vieira Machado Henriques e de Inês Vitorino

Sampaio, traz instigantes debates e reflexões sobre a discriminação algorítmica a partir da ótica dos direitos das crianças no ambiente digital, sobretudo em relação ao recorte de gênero, raça e vulnerabilidade econômica. As possíveis soluções para o problema identificadas pelo trabalho são ancoradas em uma discussão sobre direitos fundamentais, especialmente o direito à inclusão e à não discriminação e a sua incidência na infância, e na análise da relação entre os princípios éticos associados à inteligência artificial e os princípios e conceitos da proteção de dados pessoais.

O trabalho "A inteligência artificial no contexto atual: uma análise à luz das neurociências voltada para uma proposta de emolduramento ético e jurídico", de Gabrielle Sarlet, também se volta para o exame de riscos associados às novas tecnologias. A autora propõe-se a examinar o uso irreflexivo ou abusivo de tecnologias como a IA para, a partir de uma visão das neurociências, apontar o novo sentido e o alcance das vulnerabilidades na sociedade informacional. A partir da investigação dos trajetos neurocientíficos voltados para o estudo da emocionalidade, a autora identifica gatilhos e riscos associados às novas tecnologias e explora algumas possibilidades de regulamentação capazes de reordenar a relação homem-máquina, de modo que esta possa se situar em uma harmonia ancorada em uma tessitura forjada por padrões éticos que reforçam o elemento humano como primordial.

Dois artigos tratam de aspectos concretos da aplicação de sistema de IA pelo Poder Judiciário. Um deles, o artigo "O uso da inteligência artificial na repercussão geral: desafios teóricos e éticos", de autoria de Fausto Santos de Morais, traz uma contribuição bastante concreta ao debate, ao investigar a experiência do Supremo Tribunal Federal no desenvolvimento e uso do programa "Victor" para identificação de recursos e classificação de temas de repercussão geral. Dentre as diversas questões teóricas e éticas daí decorrentes, o artigo dedica atenção aos riscos de hipernormatização artificial com a sobrepadronização fática ou normativa, assim como à questão epistêmica sobre a automatização na aplicação da repercussão geral, sustentando a necessidade de ações práticas e normativas para contornar os problemas identificados.

Também em relação ao uso de IA no Poder Judiciário, o artigo "'Sob o controle do usuário': formação dos juízes brasileiros para o uso ético da IA no Judiciário", de Eunice Prado, Luciane Corrêa Münch e Márcia Corrêa Ughini Villarroel, parte da consideração da concreta e crescente implementação de sistemas de IA pelo Poder Judiciário e do necessário controle desses sistemas pelo Magistrado, que, na condição de usuário dos referidos sistemas, é identificado como "juiz-usuário". A partir de pesquisa

empírica realizada com Magistrados e de considerações sobre a potencial proficiência destes no controle dos sistemas de IA, o artigo identifica a pertinência de aprimoramento de capacidades específicas destes usuários no uso de IA.

Dando sequência, um artigo do Dossiê é dedicado a explorar o tema da explicabilidade das tecnologias de IA, um debate imprescindível em um contexto em que a opacidade de sistemas decisórios baseados em IA vem sendo progressivamente problematizada. O artigo "Da 'caixa-preta' à 'caixa de vidro': o uso da *explainable artificial intelligence* (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos", de Marco Antônio Sousa Alves e Otávio Morato de Andrade, também traz contribuições sobre a questão da explicabilidade, centrando suas atenções sobre a XAI (*explainable artificial intelligence*). Fazendo recurso às metáforas da "caixa preta" e da "caixa de vidro", o artigo investiga o potencial da XAI para explicar decisões de modelos algorítmicos e combater o enviesamento dos sistemas de IA, sugerindo que a capacidade de "se explicar" seja um requisito para a adoção de sistemas de IA em searas mais sensíveis, como, por exemplo, o auxílio à tomada de decisão judicial.

Os três últimos artigos do Dossiê exploram o papel do ser humano face a sistemas de IA. Neles, trata-se de abordar aspectos da interação da IA com o ser humano presentes no próprio funcionamento do sistema de IA, algo que é por vezes identificado como o "elemento humano" ou o "*human in the loop*", cuja introdução em mecanismos decisionais é debatida, por vezes, como critério para a legitimidade ou maior eficácia da decisão, além das situações nas quais o elemento humano é essencial para a própria viabilidade de um certo sistema de IA e suas consequências.

Dentro dessa temática, o Dossiê conta com um artigo de autoria dos próprios organizadores, Miriam Wimmer e Danilo Doneda, intitulado "'Falhas de IA' e a intervenção humana em decisões automatizadas: parâmetros para a legitimação pela humanização". A partir da investigação de diferentes categorias de problemas que podem advir de decisões tomadas por sistemas de IA, o artigo se propõe a explorar os elementos que podem atrair a necessidade de introdução de parâmetros humanos em processos de decisão impulsionados por sistemas de inteligência artificial, concluindo que, em determinados casos, a necessidade de intervenção humana pode ser identificada não apenas com base em critérios de eficiência, mas também pode se constituir em um componente ético em si mesmo. Ao mesmo tempo, o artigo busca identificar determinados parâmetros capazes de mitigar os riscos associados ao "déficit

de humanidade" e, assim, proporcionar que a intervenção humana seja modulada em diferentes níveis de intensidade, mantendo-se o atendimento aos requisitos éticos de decisões legítimas, confiáveis, justas e cognoscíveis, por seres humanos, em seus principais elementos.

Já o artigo "Microtrabalho e inteligência artificial: desafios à fruição da dignidade humana em meio à aprendizagem de máquina", de Denise Pires Fincato e Julise Carolina Lemonje, aborda a relação entre a IA e as relações de trabalho sob uma ótica pouco explorada. O trabalho lança luz sobre o tema das plataformas de microtrabalho que ofertam serviços para o desenvolvimento e aperfeiçoamento da inteligência artificial e tece considerações quanto aos seus impactos sobre a concretização do princípio e o valor dignidade humana nesse contexto. O artigo chama, ainda, atenção para os riscos associados à instrumentalização e restrições de autonomia dos trabalhadores, que se veem dissociados do valor social do seu trabalho e desconhecem elementos essenciais da atividade desempenhada.

Por fim, o artigo "Os sistemas de armas autônomas e o direito internacional: uma análise da guerra e das implicações do uso da inteligência artificial", de Gilmar Antonio Bedin, Aline Michele Pedron Leves e Laura Mallmann Marcht, aborda as significativas transformações da guerra no século XXI, com o desenvolvimento de tecnologias bélicas revolucionárias, e reflete sobre o grande desafio que a sociedade e o Direito Internacional terão para garantir a manutenção da paz mundial e evitar as significativas consequências que a utilização dos sistemas de armas autônomas poderia gerar, concluindo pela necessidade de um novo tratado internacional sobre o tema, baseado em princípios éticos e humanitários.

Assim, é com muita alegria que trazemos a público o presente Dossiê Temático, que, a partir de variados enfoques, evidencia aspectos importantes do debate contemporâneo sobre o tema. Gostaríamos de agradecer à equipe da *Revista Direito Público* pelo apoio ao longo do processo de editoração. Agradecemos, também, a todas as pessoas que submeteram seus trabalhos e aos pareceristas que, de maneira criteriosa, apresentaram críticas e sugestões quanto aos excelentes trabalhos recebidos.

Desejamos que a leitura do Dossiê contribua para a necessária reflexão sobre o tema e que suscite novas inquietações e debates!

Miriam Wimmer e Danilo Doneda

Brasília e Curitiba, janeiro de 2021.

A *Revista Direito Público* – publicação oficial do Programa de Pós-Graduação *Stricto Sensu* em Direito Constitucional (Mestrado e Doutorado Acadêmico) do Instituto Brasiliense de Direito Público (IDP) – traz aos seus leitores e leitoras o Dossiê Temático "Privacidade e Proteção de Dados Pessoais na Segurança Pública e no Processo Penal", sob a coordenação das Professoras Laura Schertel Mendes e Jacqueline Abreu, que assinam conjuntamente esta Carta.

Este Dossiê busca contribuir para o debate teórico acerca dos limites jurídicos e do modelo regulatório pertinente para práticas invasivas da privacidade ou baseadas no tratamento de dados pessoais empregadas no contexto de atividades estatais voltadas à segurança pública e investigações criminais.

Considerando o recente debate proposto no Brasil acerca da elaboração e propositura de um projeto de lei específico sobre o tema e da efetiva apresentação de um anteprojeto de lei por uma Comissão de Juristas, o tema é de relevância proeminente. Isso porque, apesar de o Anteprojeto da Lei Geral de Proteção de Dados para Segurança Pública e Investigações Criminais espelhar, em grande parte, a estrutura da Lei nº 13.709/2018 (Lei Geral de Proteção de Dados – "LGPD"), o Anteprojeto possui diferenças e especificidades em relação à LGPD, na medida em que a própria LGPD deixou em aberto a regulação específica do tratamento de dados para segurança pública e investigações criminais.

No bojo das discussões sobre proteção de dados no contexto criminal, em geral, o paradigma que vem à tona é o contraste entre sigilo e publicidade de dados em investigações. Isso porque é comum que se trate de medidas de "quebra de sigilo", pelas quais autoridades buscam acesso a dados a princípio protegidos por sigilo, como, por exemplo, dados bancários, telemáticos, telefônicos, biométricos etc.

Ocorre que o antigo paradigma "sigilo/publicidade" não é passível de abarcar e regular a infinidade de medidas e de operações de tratamento de dados pessoais que se colocam hoje à disposição para investigações criminais e para políticas de segurança pública de forma geral. Isso porque, no contexto atual, tende-se a expandir o ramo da inteligência policial, a criação e integração de bancos de dados, o uso de tecnologias de inteligência artificial alimentadas por grandes volumes de dados, isto é, buscando-se usufruir dos benefícios e do poder da *big data*.

Podemos pensar em exemplos como: (i) uso de câmeras de vigilâncias com aplicações de reconhecimento facial, leitores automáticos de placas de veículos, câmeras agregadas aos uniformes de policiais e viaturas que acompanham as suas operações; (ii) formação de bancos de dados de material genético, tal como o que foi introduzido na Lei de Execuções Penais (Lei nº 7.210, de 11 de julho de 1984), alterada pela Lei Anticrime (Lei nº 13.964, de 24 de dezembro de 2019); (iii) atividades do Conselho de Controle de Atividades Financeiras – COAF, que agregam e analisam grandes volumes de dados, para emitir relatórios de inteligência; (iv) extração de dados de redes sociais e exigências de compartilhamentos de conjuntos de dados; (v) integração de sistemas e bases de dados informáticas de segurança pública, com informações sobre boletins de ocorrência, mandados de prisão, desaparecidos, portadores de armas etc.

Toda essa atividade policial da era da *big data* envolve coleta e uso de dados de forma contínua, cumulativa, de um grande conjunto de pessoas, independentemente de possuírem histórico de condenação ou não. São informações usadas para abrir linhas de investigação, fazer provas pré-constituídas, correlacionar informações, detectar crimes e outras ações que podem ocasionar os mais diversos tipos de abusos e ilegalidades no tratamento de dados. Tais operações de tratamento podem envolver usos indevidos, usos abusivos, usos secundários (diferentes e estranhos à finalidade original), vazamentos; discriminação; erros por conta de dados sem acurácia ou desatualizados.

Nesse contexto, é inevitável e imprescindível a aplicação do direito à proteção de dados pessoais também nas esferas investigativas e da segurança pública, como já se fez com a elaboração de uma lei para o setor privado. Além da conhecida importância da proteção de dados pessoais em geral, no âmbito das investigações criminais e da segurança pública, a existência de uma lacuna legislativa sobre o tratamento de dados potencializa a exposição dos titulares de dados pessoais a violações e restrições em liberdades civis fundamentais graves, podendo levar as pessoas a serem indevidamente inquiridas e até mesmo presas.

Não são, portanto, problemas passíveis de serem endereçados com nossas ferramentas antigas. É necessário reforçar a regulação dessas atividades policiais tanto para que seja oferecida proteção adequada para o titular dos dados quanto para que as autoridades tenham segurança jurídica no tratamento desses dados pessoais. Ademais, a ausência de um marco normativo adequado para as operações de tratamento de dados pessoais nas

investigações criminais e na segurança pública também gera prejuízo significativo no bojo das cooperações internacionais, impedindo que os órgãos possam acessar tecnologias mais modernas e bases de dados com informações relevantes.

Não à toa, o Anteprojeto se inspira em diplomas internacionais, como a Diretiva nº 680/2016 da União Europeia, e em propostas norte-americanas, como as leis do estado de Washington, de Nova York e de São Francisco, que apontam para a necessidade de conferir transparência e controle institucional no tratamento de dados pelas autoridades de segurança e investigação.

O recorte existente na LGPD e o espaço para a aprovação de uma LGPD penal não significam que a noção de proteção de dados como a conhecemos hoje não seria aplicável no processo penal e na segurança pública. Dois pontos demonstram o contrário: primeiro, o reconhecimento da proteção de dados pessoais como direito fundamental autônomo pelo STF e a sua aplicação em contextos envolvendo inteligência nacional e segurança pública; segundo, a observância dos princípios e direitos e da proporcionalidade e devido processo legal, conforme determinado pela própria LGPD – que, inclusive, já determina a elaboração de relatório de impacto à proteção de dados pessoais, assim como regula em parte a cooperação internacional nessa seara.

Nesse sentido, a importância de se aprovar uma lei específica de proteção de dados para investigação criminal e segurança pública no Brasil consiste em: criar parâmetros procedimentais para que existam balizas legais concretas no tratamento de dados pelos órgãos de investigação; ampliar a transparência sobre como dados são tratados para permitir a responsabilização e desencorajar usos secundários inadvertidos; e minimizar riscos de tratamentos abusivos, ilegais e de incidentes de segurança, dentre outros.

À luz da sensibilização da discussão sobre o impacto de novas tecnologias, intersecção entre privacidade, proteção de dados e segurança pública e de novas modalidades de atividades criminosas e meios de obtenção de provas no processo penal, a academia brasileira possui importante papel na promoção de debates e no desenvolvimento de doutrina passível de endereçar problemas complexos e multidisciplinares.

Por ser um campo de pesquisa ainda novo na academia brasileira, os artigos selecionados são manifestação de esforços bem-sucedidos de avanço nessa área de pesquisa e debate. As discussões deste Dossiê perpassam

por importantes temas, como deveres de confidencialidade de informações médicas – em que se apresentam resultados de pesquisa empírica sobre "delações" de aborto autoprovocado feitas por profissionais de saúde; mecanismos de contraditório em procedimentos sigilosos de investigação que impliquem quebras de sigilo – enfrentando a difícil questão de como garantir que a narrativa investigativa não passe sem contraponto ainda em etapa anterior a (irreversíveis) afastamentos da privacidade; incidência de parâmetros de proteção de dados pessoais a uso de dados pessoais pelo COAF – dimensão de análise que joga luz aos mais diversos princípios que devem reger e ser incorporados também a esse tipo de atividade de combate a condutas criminosas na prática jurídica nacional; acesso a dados de celulares por autoridades policiais, com revisão da postura dos Tribunais Superiores sobre o tema; análise crítica de mecanismos de policiamento preditivo, tendo em vista que podem reproduzir e amplificar graves problemas de seletividade do sistema penal; uso de dados pessoais no controle migratório – atividade cada vez mais intensa no Brasil.

Há ainda artigos seminais para a área: o primeiro, do Professor Wolfgang Hoffmann-Riem, que contextualiza e fundamenta o direito à confidencialidade e integridade dos sistemas informáticos – de grande relevância para se pensar em usos de tecnologias ocultas de monitoramento; o segundo, dos Professores Paul de Hert e Serge Gutwirth, em que há a associação de noções de privacidade e proteção de dados a mecanismos de *rule of law* e a estratégias de estruturação do poder do Estado por opacidade e transparência; outro, do Professor Wolfgang Hoffmann-Riem, que contextualiza e fundamenta o direito à confidencialidade e integridade dos sistemas informáticos – de grande relevância para se pensar em usos de tecnologias ocultas de monitoramento.

Esperamos que as valorosas contribuições das autoras e dos autores possam auxiliar no diálogo sobre o impacto e a preservação da privacidade e da proteção de dados pessoais no campo das investigações criminais e das políticas de segurança pública.

Boa leitura a todos e todas!

Laura Schertel Mendes

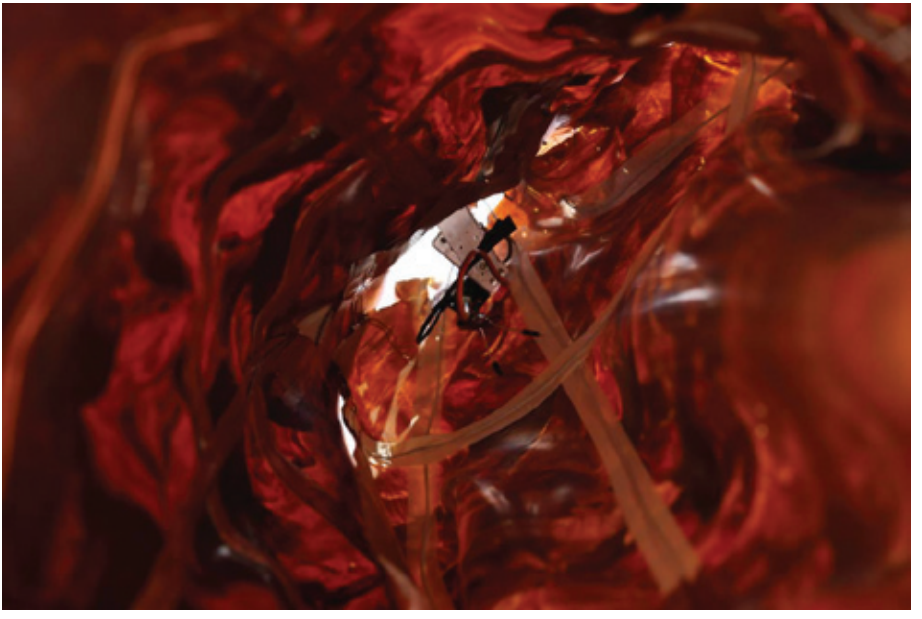
Coordenadora do Dossiê "Privacidade e Proteção de Dados Pessoais na Segurança Pública e no Processo Penal"

Jacqueline Abreu

Coordenadora do Dossiê "Privacidade e Proteção de Dados Pessoais na Segurança Pública e no Processo Penal"

Assunto Especial

Dossiê – Inteligência Artificial, Ética e Epistemologia



Katia Wille

Electronic Democracy and Digital Justice: Driving Principles for AI Regulation in the Prism of Human Rights

ALESSANDRO MANTELERO[1]
Polytechnic University of Turin (Italy).

ABSTRACT: A growing debate in several European fora is paving the way for future rules for Artificial Intelligence (AI). A principles-based approach prevails, with various lists of principles drawn up in recent years. These lists, which are often built on human rights, are only a starting point for a future regulation. It is now necessary to move forward, turning abstract principles into a context-based response to the challenges of AI. This article therefore places the principles and operational rules of the current European and international human rights framework in the context of AI applications in two core, and little explored, areas of digital transformation: electronic democracy and digital justice. Several binding and non-binding legal instruments are available for each of these areas, but they were adopted in a pre-AI era, which affects their effectiveness in providing an adequate and specific response to the challenges of AI. Although the existing guiding principles remain valid, their application should therefore be reconsidered in the light of the social and technical changes induced by AI. To contribute to the ongoing debate on future AI regulation, this article outlines a contextualised application of the principles governing e-democracy and digital justice in view of current and future AI applications.

KEYWORDS: Human Rights; Artificial Intelligence; Electronic Democracy; Digital Justice; Regulation

1 AI CHALLENGES AND HUMAN RIGHTS

Artificial Intelligence (AI) is part of our daily life. It is used to moderate public debate, fashion the social environment and support human decision-makers in various fields, including justice. AI is therefore a component of many decision-making processes affecting individuals and groups, actively

1 Orcid: 0000-0001-6020-0571.

shaping our communities and personal lives[2]. This means that AI is no longer a mere technical or marketing trend but a regulatory issue[3], given the social consequences and, in some cases, legal effects.

To correctly frame this debate, it is important to keep in mind the difference between natural and artificial intelligence, where the latter is nothing more than a data-driven and mathematical form of information processing[4]. AI is not able to think, elaborate concepts or develop theories of causality: AI merely takes a path recognition approach to order huge amounts of data and infer new information and correlations.

Data dependence is both the strength and the weakness of these systems. Poor data undermines the quality of their results[5], datafication can only partially represent reality[6] and incredibly large datasets and complex AI solutions often do not allow human decision makers to inspect and check the 'reasoning' of the machine[7]. The upshot of these technical and structural constraints can be summed up under three main headings: bias, obscurity, and ownership.

Regarding bias, the design and development of AI tools can be affected by different biases that, in many cases, differ from human bias[8]. Bias does not only concern the much debated data quality (for example selection bias)[9], but also the methodologies adopted (*e.g.*, pre-processing and data cleaning biases, measurement bias, bias in survey methodologies)[10], the target of investigation (*e.g.*, historical bias in pre-existing data-sets and under- or over-representation of certain groups in new data-sets), and the psychological attitude of the data scientists (*e.g.*, confirmation bias).

This brief listing of potential biases also reveals the human component of AI solutions, often underestimated in a misleading comparison between

2 For an analysis of the different impacts of AI on individuals and society see Council of Europe, 2018c; MANTELERO-ESPOSITO, 2021; ZUIDERVEEN BORGESIUS, 2020.

3 See European Commission, 2021; Council of Europe – Ad hoc Committee on Artificial Intelligence (CAHAI), 2020; Council of Europe, 2020b; Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), 2019; OECD, 2019; UNESCO, 2021. See also VERONESE-NUNES LOPES ESPIÑEIRA LEMOS, 2021.

4 See HILDEBRANDT, 2021.

5 See European Union Agency for Fundamental Rights, 2019.

6 See AGRE, 1994; HILDEBRANDT, 2019.

7 See KOLKMAN, 2020.

8 See CUMMINGS et al., 2018, 2; CARUANA et al., 2015; EYKHOLT et al., 2018.

9 See AI Now Institute, 2017, 4 and 16-17.

10 See VEALE-BINNS, 2017.

humans and machines. This dichotomy understates the role of human intervention in AI data processing[11] and the intentional or unintentional transposition of developers' views into the AI reference values used for classification[12].

As for obscurity, this concerns both the AI tools used and the way they impact on individuals, whose circumstances are analysed and represented through them. Not only is the way some AI applications actually function and process information unknown[13], even to data scientists, but individuals are often unaware of their being dynamically grouped on the basis of unseen correlations and inferences, without being able to know the identity of the other members of the group. Obscurity therefore entails two different consequences: first, data scientists are unable to clearly justify the specific decisions suggested by AI; and second, people are passively scrutinised by AI without having a meaningful or effective role in AI design or the opportunity to voice their collective interests[14].

This level of obscurity and the limitations to democratic participation in AI development is heightened by a third feature of many AI products: ownership. The proprietary nature of the algorithms used and, in certain cases, of the data silos used to train and implement them mean that intellectual property rights are a further barrier to access to the architecture of these applications and to public oversight[15].

These three inherent constraints – bias, obscurity, and ownership – have a direct impact on the challenges of AI and its social acceptance in monitoring and governing human activities (*e.g.*, smart cities),[16] offering personalised services (*e.g.*, predictive medicine)[17] and, more in general, supporting humans in the decision-making process.

Issues surrounding data-intensive solutions and their use in decision--making processes concern a variety of interests related to human rights and freedoms[18]. To address the growing concern about the potential impact of

11 See TUBARO-CASILLI-COVILLE, 2020; CRAWFORD-JOLER, 2018; LOIDEAIN-ADAMS, 2020.
12 See also WEST-WHITTAKER-CRAWFORD, 2019.
13 See SELBST, 163; BURRELL, 2016; BRAUNEIS-GOODMAN, 131.
14 See also GRABER, 2020; MANTELERO, 2016.
15 See PASQUALE, 2015, 193.
16 See also Privacy International, 2017; GOODMAN-POWLES, 2019. See also COHEN, 2019, 62-3.
17 See FERRYMAN-PITCAN, 2018.
18 See MANTELERO-ESPOSITO, 2021; Council of Europe, 2018c.

AI on human rights and freedoms, several initiatives have been proposed at local, national and international levels, and a variety of guidelines have been drawn up by NGOs, research centres and corporate entities. Several proposals have focused on ethics[19], often blurring the line between law and ethics, describing human rights and freedoms as ethical values with their 'ethicisation' and relativization.

This emphasis on the ethical dimension can entail the risk of extending to the field of data science an ethical imperialism whose effects are already known in biomedicine and the social sciences[20]. In this regard, previous experience in ethical assessment of scientific research suggests that careful consideration should be given to the distinction between ethical and legal values and the differences between ethical approaches[21]. Several documents providing guidelines on AI refer to ethics in a fairly broad and indefinite manner, with no clarification (or justification) of the ethical framework used[22].

Ethical responses to uncertainty in a rapidly changing technological and social environment may paradoxically become a new source of ambiguity. Discretionary and, in some cases, interest-based values risk weakening the legal framework or indirectly redefining it without following an appropriate procedure as required by the regulatory process[23].

Without underestimating the role of ethics in technology development, these considerations suggest a more balanced integration of law and ethics in AI regulation, based on the emphasis on the role of human rights as the universal cornerstone of the future architecture of AI regulation. From a regulatory perspective, the main challenge is to contextualise the legal principles and provisions enshrined in international human rights instruments, drafted in a pre-AI era, within the current scenario where predictive policing tools, automated digital propaganda and other new AI-based applications are reshaping many aspects of our society and human relations.

Regulatory initiatives have been proposed in several countries[24], many of them referring explicitly to all or some human rights. However, these are

19 See JOBIN-IENCA-VAYENA, 2019; HAGENDORFF, 2020.
20 See SCHRAG, 2017.
21 See HILDEBRANDT, 2021.
22 See RAAB, 2020. See also Independent High-Level Group on Artificial Intelligence, 2019.
23 See NEMITZ, 2018.
24 See GESLEY, 2019; Council of Europe, 2020a. See also MANHEIM-KAPLAN, 2019, 160.

often generic statements without a proper contextualisation of the rights and freedoms considered. Although it is relatively easy to agree on a general list of rights and freedoms that should underpin AI development, these lists do little to advance the regulatory process, since general principles, such as transparency or participation, can be interpreted in many different ways.

An effective contribution to the human rights debate in this field can therefore only come from a proper contextualisation of these guiding principles within the AI scenario. This means placing such rules, including the operational ones, in the context of the changes to society produced by AI and providing a more refined and specific formulation of the guiding principles with a view to possible future AI regulation.

This contextualisation of the guiding principles and rules can provide a more refined and elaborate formulation, taking into account the specific nature of AI products and services, and helping to better address the challenges arising from AI.

From a methodological perspective, an analysis of international legally binding instruments is the obligatory starting point in defining the existing legal framework, identifying its guiding values and verifying whether this framework and its principles properly address the issues raised by AI, with a view to preserving the harmonisation of the existing legal framework in the fields of democracy and justice.

The methodology adopted is therefore necessarily deductive, extracting the guiding principles from the variety of regulations concerning the fields in question. The theoretical basis of this approach relies on the assumption that the general principles provided by international human rights instruments should underpin all human activities, including AI-based innovation[25].

These guiding principles should be considered within the scenario of AI-driven transformation, which in many cases requires adaptation. They remain valid, but their implementation must be reconsidered in the light of the social and technical changes brought about by AI. This will deliver a more contextualised and granular application of these principles so that they can make a concrete contribution to the shape of future AI regulation.

25 See Council of Europe, 2020b.

Against this background, the following sections examine two critical areas of AI application: electronic democracy and digital justice. While in other areas, such as data protection and biomedicine, the specific nature of the sectors and recent soft-law regulatory initiatives[26] make it possible to draft some provisions for future AI regulation[27], in these two realms this is much more difficult. In addition, key principles that can be seen as guiding elements of future AI regulation, such as transparency and explainability[28], are open to varying interpretations and implementations, given the higher political significance of both democracy and justice. The analysis therefore focuses on high-level principles and their contextualisation, resulting in a more limited elaboration of key guiding provisions.

2 AI AND ELECTRONIC DEMOCRACY

Democracy covers an extremely wide array of societal and legal issues[29], most of them likely to be implemented with the support of ICT[30]. In this scenario, AI can play an important role in the present and future development of digital democracy in an information society.

The broad dimension of this topic makes it difficult to identify a single binding sector-specific legal instrument for reference. Several international instruments deal with democracy and its different aspects, starting with the UN Declaration of Human Rights and the International Covenant on Civil and Political Rights. Similarly, in the European context, key principles for democracy are present in several international sources.

Based on Article 25 ICCPR, we can identify two main areas of intervention related to electronic democracy: (i) participation[31] and good governance, and (ii) elections. Undoubtedly, it is difficult or impossible to draw a red line between these fields as they are interconnected in various ways. AI can have an impact on all of them: participation (*e.g.*, citizens engagement, participation platforms), good governance (*e.g.*, e-government,

26 See for example Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data, 2019; CEPEJ, 2018.

27 See MANTELERO, 2020.

28 *E.g.*, SELBST-BAROCAS, 2018; EDWARDS-VEALE, 2017; DIAKOPOULPS, 2013. See also KAMINSKI-MALGIERI, 2020.

29 *E.g.*, Council of Europe, Directorate General of Democracy – European Committee on Democracy and Governance, 2016.

30 *E.g.*, Council of Europe Directorate General of Democracy and Political Affairs and Directorate of Democratic Institutions, 2009; Council of Europe, 2009a, Article 2.2.iii.

31 For a more detailed analysis see Faye Jacobsen, 2013. See also MAISLEY, 2017.

decision-making processes, smart cities), pre-electoral phase (*e.g.*, financing, targeting and profiling, propaganda), elections (*e.g.*, prediction of election results, e-voting), and the post-election period (*e.g.*, electoral dispute resolution).

As in any classification, this distinction is characterised by a margin of directionality. It is worth pointing here out that this is a functional classification based on different AI impacts, with no intention to provide a legal or political representation of democracy and its different key elements. The relationship between participation, good governance, and elections can therefore be considered from different angles and shaped in different ways, unifying certain areas or further subdividing them.

Participation is expressed both through taking part in the democratic debate and through the electoral process, but the way that AI tools interact with participation in these two cases differs and there are distinct international legal instruments specific to the electoral process.

2.1 Participation 2and good governance

The right to participate in public affairs (Article 25 Covenant) is based on a broad concept of public affairs[32], which includes public debate and dialogue between citizens and their representatives, with a close link to freedom of expression, assembly, and association[33]. In this respect, AI is relevant from two different perspectives: as a means to participation and as the subject of participatory decisions.

Considering AI as a means, technical and educational barriers can undermine the exercise of the right to participate. Participation tools based on AI should therefore consider the risks of under-representation and lack of transparency in participative processes (for example platforms for the drafting of bills). At the same time, AI is also the subject of participatory decisions, as they include decisions on the development of AI in general and its use in public affairs.

AI-based participative platforms (*e.g.*, Consul, Citizenlab, Decidim[34]) can make a significant contribution to the democratic process, facilitating

32 See UN Human Rights Committee, 1996.

33 See also UN Committee on Economic, Social and Cultural Rights, 1981, para 5.

34 Information on these platforms is available at https://decidim.org/; https://consulproject.org/en/; https://www.citizenlab.co/. Accessed: 29 dec. 2019.

citizen interaction, prioritising of objectives, and collaborative approaches in decision-making[35] on topics of general interests at different levels (neighbourhood, municipality, metropolitan area, region, country)[36].

Specific issues arise in relation to AI tools for democratic participation (including those for preventing and fighting corruption[37]), which are associated with the following four main areas: transparency, accountability, inclusiveness, and openness. In this regard, the general principles set out in international binding instruments have an important implementation in the Recommendation CM/Rec(2009)1 of the Committee of Ministers of the Council of Europe to member states on electronic democracy (e-democracy), which provides a basis for further elaboration of the guiding principles in the field of AI with regard to democracy.

Transparency is a requirement for the use of technological applications for democratic purposes[38]. This principle is common to other fields, such as healthcare[39], but is a context-based notion. While in healthcare transparency is closely related to self-determination, here it is not only a requirement for citizens' self-determination with respect to a technical tool but is also a component of the democratic participatory process[40]. Transparency no longer has an individual dimension but assumes a collective dimension as a guarantee of the democratic process.

In this context, the use of AI-based solutions for e-democracy must be transparent in respect of their logic and functioning (*e.g.*, content selection in participatory platforms) providing clear, easily accessible, intelligible, and updated information about the AI tools used and their justification[41].

Moreover, the implementation of this notion of transparency should also consider the range of different users of these tools, adopting an accessible approach[42] from the early stages of the design of AI tools. This is to ensure effective transparency with regard to vulnerable and impaired groups, giving added value to accessibility in this context.

35 See also Council of Europe, 2017a.

36 See also Council of Europe, 2009c.

37 See United Nations, Convention against Corruption, 2003, Article 13.

38 See Council of Europe, 2009b, para 6.

39 See Council of Europe, 1997.

40 See also Council of Europe, 2017a.

41 See Council of Europe, 2009b, para 6 and Appendix, para P.57. See also Council of Europe, 2016b, Appendix, paras 2.1.3 and 3.2. On the importance of justification see Hildebrandt, 2018b, 271-3.

42 See also Council of Europe, 2018b, Appendix, para B.IV.

Transparency and accessibility are closely related to the nature of the architecture used to build AI systems. Open source and open standards can therefore contribute to democratic oversight of the most critical AI applications[43]. There are cases where openness is affected by limitations, due to the nature of the specific AI application (for example crime prevention). In these cases, auditability, as well as certification schemes, play a more important role than they already do in relation to AI systems in general[44].

In the context of AI applications to foster democratic participation, an important role can be also played by interoperability[45] as it facilitates integration between different services/platforms for e-democracy and at different geographical levels. This aspect is already relevant for e-democracy in general[46], and should therefore be extended to the design of AI-based systems.

Another key principle in e-democracy is accountability. In this regard, to be accountable, AI service providers and entities using AI-based solutions for e-democracy shall adopt forms of algorithm vigilance that promote the accountability of all relevant stakeholders by assessing and documenting the expected impacts on individuals and society in each phase of the AI system lifecycle on a continuous basis, to ensure compliance with human rights, the rule of law and democracy[47].

Finally, given the role of media in the context of democratic participation[48], AI applications must not compromise the confidentiality and security of communications and protection of journalistic sources and whistle-blowers[49].

In addressing the different aspects of developing AI solutions for democratic participation, a first consideration is that a democratic approach is incompatible with a techno-determinist approach. AI solutions to address societal problems should therefore be the result of an inclusive process. Hence, legal values such as the protection of minorities, pluralism and

43 See also Council of Europe, 2009b, para 6 and Appendix, paras G.58 and P.54.

44 It is worth to underline that auditing and certification schemes play an important role also in cases of open-source AI architecture, as this nature does not imply per se absence of bias or any other shortcomings. See also Council of Europe, 2009b, Appendix, paras P. 55 and G.57.

45 See also Council of Europe, 2009b, Appendix, paras P. 56, G.56, 59 and 60.

46 See also Council of Europe, 2009b, para 6.

47 See also Council of Europe, 2020b.

48 See Council of Europe, 2016a, Appendix, para 2; Council of Europe Parliamentary Assembly, 2019a.

49 See also Council of Europe Parliamentary Assembly, 2019b; Council of Europe, 2014.

diversity should be a necessary consideration in the development of these solutions.

From a democratic perspective, the first question we should ask is: do we really need an AI-based solution to a given problem as opposed to other options[50], considering the potential impact of AI on rights and freedoms? If the answer to this question is yes, the next step is to examine value-embedding in AI development[51].

The proposed AI solutions must be designed from a human rights-oriented perspective, ensuring full respect for human rights and fundamental freedoms, including the adoption of assessment tools and procedures for this purpose[52]. In the case of AI applications with a high impact on human rights and freedoms, such as electoral processes, legal compliance should be prior assessed. In addition, AI systems for public tasks should be auditable and, where not excluded by competing prevailing interests, audits should be publicly available.

Another important aspect to be considered is the public-private partnership that frequently characterises AI services for citizens[53], weighing which is the best choice between in-house and third-party solutions, including the many different combinations of these two extremes. In this regard, when AI solutions are fully or partially developed by private companies, transparency of contracts and clear rules on access and use of citizens' data have a critical value in terms of democratic oversight.

Restrictions on access and use of citizens' data are not only relevant from a data protection perspective (principles of data minimisation and purpose limitation) but more generally with regard to the bulk of data generated by a community, which also includes non-personal data and aggregated data. This issue should be considered as a component of democracy in the digital environment, where the collective dimension of the digital resources generated by a community should entail forms of citizen control and oversight, as happens for the other resources of a territory/community.

50 See also Council of Europe, 2020b.

51 See also Council of Europe, 2019a, para 7.

52 See Council of Europe, 2009b, paras 5 and 6, and Appendix, para G.67. See also Mantelero, 2018.

53 See MIKHAYLOV-ESTEVE-CAMPION, 2018.

The considerations already expressed above on openness as a key element of democratic participation tools should be recalled here, given their impact on the design of AI systems. Furthermore, the design, development and deployment of these systems should also consider the adoption of an environmentally friendly and sustainable strategy[54].

Finally, it is worth noting that while AI-design is a key component of these systems, design is not neutral. Values can be embedded in technological artefacts[55], including AI systems. These values can be chosen intentionally and, in the context of e-democracy, this must be based on a democratic process. But they may also be unintentionally embedded into AI solutions, due to the cultural, social and gender composition of AI developer teams. For this reason, inclusiveness has an added value here, in terms of inclusion and diversity[56] in AI development.

The principles discussed for e-democracy can be repeated with regard to good governance[57]. This is the case with smart cities and sensor-based environmental management, where open, transparent and inclusive decision-making processes play a central role[58]. Similarly, the use of AI to supervise the activities of local authorities[59], for auditing and anticorruption purposes[60], should be based on openness (open source software), transparency and auditability.

More generally, AI can be used in government/citizen interaction to automate citizen' inquiries and information requests[61]. However, in these cases, it is important to guarantee the right to know we are interacting with a machine[62] and to have a human contact point. Moreover, access to public services must not depend on the provision of data that is unnecessary and not proportionate to the purpose.

54 See also Council of Europe, 2009b, Appendix, para P. 58.

55 See also VERBEEK, 2011, 41-65.

56 See also Council of Europe, 2020b, Appendix, para 3.5.

57 See also Council of Europe, 2009b, Appendix, para P. 4; Council of Europe, 2004; Committee of Ministers of the Council of Europe, 2008.

58 See also Privacy International, 2017.

59 See also Council of Europe, 2019b, Appendix, paras 4 and 9.

60 See also SAVAGET-CHIARINI-EVANS, 2019 (discussing the Brazilian case of the 'Operação Serenata de Amor').

61 See MEHR, 2017.

62 See also Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108), 2019, para 2.11.

Special attention should also be paid to the potential use of AI in human-machine interaction to implement nudging strategies[63]. Here, due to the complexity and obscurity of the technical solutions adopted, AI can increase the passive role of citizens and negatively affect the democratic decision-making process. Otherwise, an active approach based on conscious and active participation in community goals should be preferred and better managed by AI participation tools. Where adopted, nudging strategies should still follow an evidence-based approach.

Finally, the use of AI systems in governance tasks raises challenging questions about the relationship between human decision-makers and the role of AI in the decision-making process[64]. These issues are more relevant with regard to the functions that have a high impact on individual rights and freedoms, as in the case of jurisdictional decisions[65].

2.2 Elections

The impact of AI on electoral processes is broad and concerns the pre-election, election, and post-election phases in different ways. However, an analysis focused on the stages of the electoral process does not adequately highlight the different ways in which AI solutions interact with it.

The influence of AI is therefore better represented by the following distinction: AI for the electoral process (e-voting, predictions of results, and electoral dispute resolution) and AI for electoral campaigns (micro-targeting and profiling, propaganda and fake news). While in the first area AI is mainly a technological improvement of an existing process, in the field of electoral campaigning AI-based profiling and propaganda raise new concerns that are only partially addressed by the existing legal framework. In addition, several documents have emphasised the active role of states in creating an enabling environment for freedom of expression[66].

63 On the use of nudging in the smart city context, see Ranchordás, 2019; Gandy-Nemorin, 2019. See generally Sunstein, 2015a and 2015b; Thaler-Sunstein, 2008; Sunstein-Thaler, 2003.

64 See also CITRON-CALO, 2021.

65 See Section 3.

66 See Council of Europe, 2018a; The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., 2017. See also Council of Europe, 2016b, Appendix, paras 1.5, 2.1 and 3; European Commission for Democracy trough Law (Venice Commission), 2019, para 151.E; Bukovska, 2020. See also Bychawska-Siniarska, 2017.

As regards the technological implementation of e-democracy (e-voting, prediction of results, and electoral dispute resolution), some of the key principles mentioned with regard to democratic participation are also relevant here. Accessibility, transparency, openness, risk management and accountability (including the adoption of certification and auditing procedures) are fundamental elements of the technological solutions adopted in these stages of the electoral process[67].

As regards AI for campaigning (micro-targeting and profiling, propaganda and fake news), some of the issues raised concern the processing of personal data in general. The principles set out in Convention 108+ can therefore be applied and properly contextualised[68].

More specific and new responses are needed in the case of propaganda and disinformation[69]. Here the existing binding and non-binding instruments do not set specific provisions, given the novelty of the disinformation based on new forms of communication, such as social networks, which differ from traditional media[70] and often bypass the professional mediation of the journalists.

However, general principles, such as the principle of non-interference by public authorities on media activities to influence elections[71], can be extended to these new forms of propaganda and disinformation. Considering the use of AI to automate propaganda, future AI regulation should extend the scope of the general principles of non-interference to AI-based systems used to provide false, misleading and harmful information. In addition, to prevent such interference, states[72] and social media providers should adopt a by-design approach to increase their resilience to disinformation and propaganda.

Similarly, the obligation to cover election campaigns in a fair, balanced, and impartial manner[73] should entail obligations for media and social media

67 See Council of Europe, 2017b, Appendix I, paras 1, 2, 32, and 35-40. See also Council of Europe, Directorate General of democracy and Political Affairs – Directorate of Democratic Institutions, 2011.

68 See Council of Europe, 2010; Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data, 2019.

69 See MANHEIM-KAPLAN, 2019; European Commission, Networks, Content and Technology- Directorate--General for Communication, 2018.

70 See also Council of Europe, 2011.

71 See Council of Europe, 2007, para I.1.

72 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., 'Joint Declaration on "Fake News," Disinformation and Propaganda', para 2.c.

73 See Council of Europe, 2007, para II.1.

operators regarding the transparency of the logic of the algorithms used for content selection,[74] ensuring pluralism and diversity of voices[75], including critical ones[76].

Moreover, states and intermediaries should promote and facilitate access to tools to detect disinformation and non-human agents, as well as support independent research on the impact of disinformation and projects offering fact-checking services to users[77].

Given the important role played by advertising in disinformation and propaganda, the criteria used by AI-based solutions for political advertising should be transparent[78], auditable and provide equal conditions to all the political parties and candidates[79]. In addition, intermediaries should review their advertising models to ensure that they do not adversely affect the diversity of opinions and ideas[80].

3 AI AND DIGITAL JUSTICE

As in the case of democracy, the field of justice is a broad domain and analysing the whole spectrum of the consequences of AI on justice would be too ambitious. In line with the scope of this contribution, this section sets out to describe the main challenges associated with the use of AI in digital justice and the principles which, based on international legally binding instruments, can contribute to its future regulation.

This analysis is facilitated by the European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment, adopted by the CEPEJ in 2019, which directly addresses the relationship between justice and AI. Although this non-binding instrument is classed as an ethical charter, to a large extent it concerns legal principles enshrined in international instruments.

74 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., Appendix, paras 2.1.3 and 2.3.5.

75 See also EU Code of Practice on Disinformation, 2018.

76 See also Council of Europe, 2016a, Appendix, para 15.

77 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., para 4.e; European Commission for Democracy trough Law (Venice Commission), 2019, para 151.D.

78 See also Council of Europe Parliamentary Assembly, 2019a, paras 9.2 and 11.1; European Commission for Democracy trough Law (Venice Commission), 2019, paras 151.A and 151.B.

79 See also Council of Europe, 2007, para II.5.

80 See also The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression et al., para 4.e.

Guiding principles for the development of AI in the field of digital justice can be derived from the following binding instruments: the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights, the Convention for the Protection of Human Rights and Fundamental Freedoms, the International Convention on the Elimination of All Forms of Racial Discrimination, the Convention on the Elimination of All Forms of Discrimination against Women, and the Convention for the Protection of Human Rights and Fundamental Freedoms[81].

Given the range of types and purposes of operations in this field and the various professional figures and procedures involved, this section makes a functional distinction between two areas: (i) judicial decisions and alternative dispute resolutions (ADRs) and (ii) crime prevention/prediction. Before analysing and contextualising the key principles relating to these two areas, we should offer some general observation, which may also apply to the action of the public administration as a whole[82].

First of all, it is worth noting that – compared to human decisions, and more specifically judicial decisions – the logic behind AI systems does not resemble legal reasoning. Instead, they simply execute codes based on a data-centric and mathematical/statistical approach.

In addition, error rates for AI are close to, or lower than, the human brain in fields such as image labelling, but more complicated decision--making tasks have higher error rates. This is the case with legal reasoning in problem solving[83]. At the same time, while a misclassification of an image of a cat may have limited adverse effects, an error rate in legal decisions[84] has a high impact on rights and freedom of individuals.

It is worth pointing out that the difference between errors in human and machine decision-making has an important consequence in terms of scale: while human error affects only individual cases, poor design and bias in AI inevitably affect all people in the same or similar circumstances, with AI tools being applied to a whole series of cases. This may cause group discrimination, adversely affecting individuals belonging to different traditional and non-traditional categories[85].

81 See also, with regard to the EU area, the Charter of Fundamental Rights of the European Union.

82 See Section 2.

83 See also OSOBA-WELSER, 2017, 18. See also Cummings et al., 2018, 13.

84 See Aletras et al., 2016. See also Pasquale-Cashwell, 2018; Hildebrandt, 2018a.

85 See WACHTER, 2021; MITTELSTADT, 2017, 485. See also TAYLOR-FLORIDI-VAN der Sloot, 2017.

Given the textual nature of legal documents, natural language processing (NLP) can play an important role in AI applications for the justice sphere[86]. This raises several critical issues surrounding commercial solutions developed with a focus on the English-speaking market, making them less effective in a legal environment that uses languages other than English[87]. Moreover, legal decisions are often characterised by implicit unexpressed reasoning, which may be amenable to expert systems, but not by language--based machine learning tools. Finally, the presence of general clauses requires a prior knowledge of the relevant legal interpretation and continual updates which cannot be derived from text mining.

All these constraints suggest a careful and more critical adoption of AI in the field of justice than in other domains and, with regard to court decisions and ARDs, suggest following a distinction between cases characterised by routinely and fact-based evaluations and cases characterised by a significant margin for legal reasoning and discretion[88].

3.1 ADRs and court decisions

Several so-called Legal Tech AI products do not have a direct impact on the decision-making processes in courts or alternative dispute resolutions (ADRs), but rather facilitate content and knowledge management, organisational management, and performance measurement[89]. These applications include, for example, tools for contracts categorisation, detection of divergent or incompatible contractual clauses, e-discovery, drafting assistance, law provision retrieval, assisted compliance review. In addition, some applications can provide basic problem-solving functions based on standard questions and standardised situations (*e.g.*, legal chatbots).

Although AI has an impact in such cases on legal practice and legal knowledge that raises various ethical issues[90], the potential adverse consequences for human rights, democracy and the rule of law are limited. To a large extent, they are related to inefficiencies or flaws of these systems.

86 But see OSWALD, 2018; PASQUALE-CASHWELL, 2018.
87 See Council of Bars & Law Societies of Europe, 2020, 29.
88 See the following Section on the distinction between codified justice and equitable justice.
89 See CEPEJ, 2018, Appendix II.
90 See also NUNEZ, 2017.

In the case of content and knowledge management, including research and document analysis, these flaws can generate incomplete or inaccurate representations of facts or situations, but this affects the meta-products, the results of a research tool that need to be interpreted and adequately motivated when used in court. Liability rules, in the context of product liability, for instance, can address these issues.

In addition, bias (poor case selection, misclassification etc.) affecting standard text-based computer-assisted search tools for the analysis of legislation, case-law, and literature[91], can be countered by suitable education and training of legal professionals and the transparency of AI systems (that is the description of their logic, potential bias and limitations) can reduce the negative consequences.

Transparency should also characterise the use by courts of AI for legal research and document analysis. Judges must be transparent as to which decisions depend on AI and how the results provided by AI are used to contribute to the arguments, in line with the principles of fair trial and equality of arms[92].

Finally, transparency can play an important role with regard to legal chatbots based on AI, making users aware of their logic and the resources used (for example list of cases analysed). Full transparency should also include the sources used to train these algorithms and access to the database used to provide answers. Where these databases are private, third-party audits should be available to assess the quality of datasets and how potential biases have been addressed, including the risk of under- or over-representation of certain categories (non-discrimination).

Further critical issues affect AI applications designed to automate alternative dispute resolution or to support judicial decision. Here, the distinction between codified justice and equitable justice[93] suggests that AI should be circumscribed for decision-making purposes to cases characterised by routine and fact-based evaluations. This entails the importance to carry out further research on the classification of the different kind of decisional processes to identify those routinised applications of legal reasoning that

91 See the notion of e-justice in Council of Europe, 2009b, Appendix, para 38.
92 See also CEPEJ, 2018.
93 See RE-SOLOW-NIEDERMAN, 2019, 252-4.

can be demanded to AI, preserving in any case human overview that also guarantees legal creativity of decision-makers[94].

Regarding equitable justice, as the literature points out[95], its logic is more complicated than the simple outcome of individual cases. Expressed and unexpressed values and considerations, both legal and non-legal, characterise the reasoning of the courts and are not replicable by the logic of AI. ML-based systems are not able to perform a legal reasoning. They extract inferences by identifying patterns in legal datasets, which is not the same as the elaboration of legal reasoning.

Considering the wider context of the social role of courts, jurisprudence is an evolving system, open to new societal and political issues. AI path--dependent tools could therefore stymie this evolutive process: the deductive and path-dependent nature of certain AI solutions can undermine the important role of human decision-makers in the evolution of law in practice and legal reasoning.

Moreover, at the individual level, path-dependency may also entail the risk of 'deterministic analyses'[96], prompting the resurgence of deterministic doctrines to the detriment of doctrines of individualisation of the sanction and with prejudice to the principle of rehabilitation and individualisation in sentencing.

In addition, in several cases, including ADR, both the mediation between the parties' demands and the analysis of the psychological component of human actions (fault, intentionality) require emotional intelligence that AI systems do not have.

These concerns are reflected in the existing legal framework provided by the international legal instruments. The Universal Declaration of Human Rights (Articles 7 and 10), the International Covenant on Civil and Political Rights (Article 14), the Convention for the Protection of Human Rights and Fundamental Freedoms (Article 6) and also the Charter of Fundamental Rights of the European Union (Article 47) stress the following key requirements with regard to the exercise of judicial power: equal treatment before the law,

94 See also Clay, 2019), 58. In this regard, for example, a legal system that provides compensation for physical injuries on the basis of the effective patrimonial damages could be automatised, but it will not be able to reconsider the foundation of the legal reasoning and extend compensation to non-personal and existential damages.

95 See RE-SOLOW-NIEDERMAN, 2019.

96 See CEPEJ, 2018, 9.

impartiality, independence and competency. AI tools do not possess these qualities, and this limits their contribution to the decision-making process as carried out by courts.

As stated by the European Commission for the Efficiency of Justice, 'the neutrality of algorithms is a myth, as their creators consciously or unintentionally transfer their own value systems into them'. Many cases of biases regarding AI applications confirm that these systems too often – albeit in many cases unintentionally – provide a partial representation of society and individual cases, which is not compatible with the principles of equal treatment before the law and non-discrimination[97]. Data quality and other forms of quality assessment (impact assessment, audits, etc.) can reduce this risk but, given the degree of potentially affected interests in the event of biased decisions, the risks remain high in the case of equitable justice and seem disproportionate to the benefits largely in terms of efficiency for the justice system[98].

Further concerns affect the principles of fair trial and of equality of arms[99], when court decisions are based on the results of proprietary algorithms whose training data and structure are not publicly available[100]. A broad notion of transparency might address these issues in relation to the use of AI in judicial decisions, but the transparency of AI – a challenging goal in itself – cannot address the other structural and functional objections cited above.

In addition, data scientists can shape AI tools in different ways in the design and training phases, so that were AI tools to become an obligatory part of the decision-making process, governments selecting the tools to be used by the courts could potentially indirectly interfere with the independence of the judges.

This risk is not eliminated by the fact that the judge remains free to disregard AI decisions, providing a specific motivation. Although human oversight is an important element[101], its effective impact may be undermined

97 See also CEPEJ, 2018.

98 See also Council of Europe, 2020b, Appendix, para 11. See also Pasquale and Cashwell, 'Prediction, Persuasion, and the Jurisprudence of Behaviourism'.

99 See also CEPEJ, 2018, Appendix I, para 138.

100 See also CEPEJ, 2018, Appendix I, para 131.

101 See also ZALNIERIUTE-BENNETT MOSES-WILLIAMS, 2019. In the case of administrative decisions, this propensity may be reinforced by the threat of potential sanctions for taking a decision that ignores results

by the psychological or utilitarian (cost-efficient) propensity of the human decision-maker to take advantage of the solution provided by AI[102].

3.2 Crime prevention

The complexity of crime detection and prevention has stimulated research in AI applications to facilitate human activities. In recent years, several solutions[103] and a growing literature have been developed in the field of predictive policing, which is a proactive data-driven approach to crime prevention. Essentially, the available solutions pursue two different goals: to predict where and when crimes might occur or to predict who might commit a crime[104].

These two purposes have a distinct potential impact on human rights and freedom, which is more pronounced when AI is used for individual predictions. However, in both cases, we can repeat here the considerations about the general challenges related to AI (obscurity, intellectual property rights, large-scale data collection[105], etc.) discussed in the previous sections and partially addressed by transparency, data quality, data protection, auditing and the other measures[106]. It is worth noting that the role of transparency in the judicial context could be limited so as not to frustrate the deterrent effect of these tools[107]. Full transparency could therefore be replaced by auditing and oversight by independent authorities.

Leaving aside the organisational aspects regarding the limitation of police officers' self-determination in the performance of their duties, the main issues with regard to the use of AI to predict crime on geographic and temporal basis concern the impact of these tools on the right to non-discrimination[108]. Self-fulfilling bias, community bias[109] and historical bias[110]

produced by analytics; Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data, 2019, para 3.4.

102 See also CITRON-CALO, 2021; MANTELERO, 2019; BRAUNEIS-GOODMAN, 2018, 127.

103 See ZAVRŠNIK, 2019; European Union Agency for Fundamental Rights, 2018, 98-100; OSOBA-WELSER, 2017.

104 For a taxonomy of predictive methods, see PERRY et al., 2013.

105 See also Council of Europe, 2001, Appendix, para 42.

106 See also RICHARDSON-SCHULTZ-CRAWFORD, 2019.

107 See also OSWALD, 2018; Barrett, 2017, 361-2.

108 See European Union Agency for Fundamental Rights, 2018, 10.

109 See also BARRETT, 358-9.

110 See BENNETT MOSES-Chan, 2018.

can produce forms of stigmatisation for certain groups and the areas where they typically live.

Where data analysis is used to classify crimes and infer evidence on criminal networks, proprietary solutions raise issues in terms of respect for the principles of fair trial and of equality of arms with regard to the collection and use of evidence. Moreover, if the daily operations of policy departments are guided by predictive software, this raises a problem of accountability of the strategies adopted, as they are partially determined by software and hence by software developer companies, rather than the police.

A sharper conflict with human rights arises in the area of predictive policing tools that use profiling to support individual forecasting. Quite apart from the question of data processing and profiling[111], these solutions can also adversely affect the principle of presumption of innocence, procedural fairness, and the right to non-discrimination[112].

While non-discrimination issues could be partially addressed, the remaining conflicts seem to be more difficult to resolve. From a human rights standpoint and in terms of proportionality (including the right to respect for private and family life)[113], the risk of prejudice to these principles seems high and not adequately countered by the evidence of benefits for individual and collective rights and freedoms[114]. In the light of future AI regulation, this should urge careful consideration of these issues, taking into account the distinction between the technical possibilities of AI solutions and their concrete benefits in safeguarding and enhancing human rights and freedoms.

Finally, from a wider and comprehensive human rights perspective, the focus on crime by data-driven AI tools drives a short-term factual approach that underrates the social issues that are often crime-related and require long-term social strategies involving the effective enhancement of individual and social rights and freedoms[115].

111 See LYNSKEY, 2019; MANTELERO-VACIAGO, 2015; Hildebrandt-Gutwirth, 2008.
112 See also Council of Europe, 2001, Appendix, paras 47 and 49.
113 See van BRAKEL-De Hert, 2011, 183.
114 See MEIJER-WESSELS, 2019.
115 See also ROSENBAUM, 2006, 245-66.

4 CONCLUSIONS

The latest wave of AI development is having a growing transformative impact on society and raises new questions in several fields, from predictive medicine and media content moderation to the quantified self and judicial systems.

With a view to preserving the harmonisation of the existing legal framework in the field of human rights, this article sets out to contribute to the debate on future AI regulation by building on existing binding instruments, contextualising their principles and providing key regulatory guidance in the fields of electronic democracy and digital justice.

This approach is based on the assumption that all human activities, including innovation through AI, should be underpinned by the general international principles on human rights. Moreover, only the human rights framework can provide a universal reference for the regulation of AI, while other yardsticks (for example ethics) do not have the same global dimension, are more context-dependent and characterised by a variety of theoretical approaches.

The findings of this analysis show that a limited number of cases do share common principles (for example individual self-determination, non-discrimination, human oversight). This is due to several factors.

First, some principles are sector specific. This is the case, for instance, with the independence of judges or the principles of fair trial and equality of arms, which concern justice alone[116].

Second, some guiding principles are shared by different areas, but with different nuances in each context. This is true for transparency, which is often regarded as pivotal in AI regulation, but takes on different meanings in different regulatory contexts.

Transparency, as a means to control the power over data in the hands of public and private entities, is crucial with regard to AI applications for democratic participation and good governance. In the context of justice, transparency has a more complex significance, being vital to safeguard fundamental rights and freedoms (*e.g.*, use of AI in the courts), but also

116 See also the principles of equitable access and of beneficence in health sector, or the principles of non-interference by public authorities in the media to influence elections and the obligation to treat all political parties and candidates equally in electoral advertising.

requiring limitation to avoid prejudicing competing interests (*e.g.*, crime detection and prevention in predictive policing).

We can therefore conclude that transparency is a guiding principle, but we must go beyond a mere claim for transparency as a key principle for AI regulation. As with other key principles (such as participation, inclusion, democratic oversight, and openness), a proper contextualisation is needed, with provisions that take into account the different contexts in which they operate.

Third, some principles are different, but belong to the same conceptual area, assuming various nuances in the different contexts. This is the case with accountability and guiding principles on risk management in general. Here the level of detail and related requirements can be more or less elaborate. While, for instance, in the field of data protection there are several provisions implementing these principles with a significant degree of detail[117], in the case of democracy and justice these principles are less developed in data--intensive applications such as AI.

Finally, there are certain components of an AI regulatory strategy that are not principles, but operational approaches and solutions, common to the different areas though requiring context-based development. This is the case with the important role played by education and training.

Such considerations suggest only partial harmonisation is achievable. The framework of future international AI regulation should therefore be based on a legally binding instrument that includes both general provisions – focusing on common principles and operational solutions – and more specific and sectoral provisions, covering those principles that are relevant only in a given field or cases where the same principle is contextualised differently in the different fields.

The analysis carried out in the previous sections has also confirmed that the existing framework based on human rights can provide an appropriate and common context for the development of more specific binding instruments to regulate AI, in line with the principles enshrined in the international legal instruments and capable of effectively addressing the issues raised by AI.

117 See Council of Europe, 2018a, and Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data, 2019.

With a view to future regulation of AI, this study does not rule out a number of gaps, largely due to the fact that in broad areas, such as democracy and justice, differing options and interpretations are available, depending on the political and societal vision of the future relationship between humans and machines. Further investigation in the field of human rights and AI, as well as the ongoing debate at international and regional level, will contribute to bridging these gaps.

REFERENCES

AGRE, P.E. Surveillance and Capture: Two Models of Privacy. *The Information Society* 10, 101,1994.

AI NOW INSTITUTE. AI Now 2017 Report. New York, 2017. Available at https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf. Accessed: 26 oct. 2017.

ALETRAS, N. et al. Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science* 2, 93, 2016, doi:10.7717/peerj-cs.93.

BARRETT, L. Reasoably Suspicious Algorithms: Predictive Policing at the United States Border. *New York University Review of Law & Social Change* 41, 327, 2017.

BENNETT MOSES L.; CHAN, J. Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability. *Policing and Society* 28, 806, 2018.

BRAUNEIS R.; GOODMAN, E.P. Algorithmic Transparency for the Smart City. *Yale J.L. & Tech.* 20, 103, 2018.

BUKOVSKA, B. Spotlight on Artificial Intelligence and Freedom of Expression #SAIFE'. Organization for Security and Co-operation in Europe, 2020. Available at https://www.osce.org/files/f/documents/9/f/456319_0.pdf. Accessed: 11 aug. 2020.

BURRELL, J. How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms. *Big Data & Society* 3(1), 2016, doi: 10.1177/2053951715622512.

BYCHAWSKA-SINIARSKA, D. Protection the Right to Freedom of Expression under the European Convention on Human Rights. Council of Europe, 2017.

CARUANA, R. et al. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015.

CEPEJ – European Commission for the Efficiency of Justice. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment, 2018.

CITRON D.K.; CALO R. The Automated Administrative State: A Crisis of Legitimacy. *Emory Law Journal* 70(4), 2021.

CLAY T. (ed), *L'arbitrage en ligne. Rapport du Club des Juristes*, 2019. Available at https://www.leclubdesjuristes.com/les-commissions/larbitrage-en-ligne/. Accessed: 30 may 2020.

COHEN, J.E. *Between Truth and Power. The Legal Construction of Informational Capitalism*. New York, Oxford University Press, 2019.

Committee of Ministers of the Council of Europe. The 12 Principles of Good Governance enshrined in the Strategy on Innovation and Good Governance at local level, 2008.

Council of Bars & Law Societies of Europe. CCBE Considerations on the Legal Aspects of Artificial Intelligence, 2020.

Council of Europe – Ad hoc Committee on Artificial Intelligence (CAHAI). Feasibility Study, CAHAI(2020)23, 2020. Available at https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da. Accessed: 29 jul.2021.

Council of Europe – Committee of the Convention for the Protection of Individuals with regards to Processing of Personal Data (Convention 108). Guidelines on artificial intelligence and data protection, 2019. Available at https://rm.coe.int/guidelines-on-artificial-intelligence-and-data-protection/168091f9d8. Accessed: 20 feb. 2020.

Council of Europe Directorate General of Democracy and Political Affairs and Directorate of Democratic Institutions, Project «Good Governance in the Information Society», CM(2009)9 Addendum 3, 2009

Council of Europe Parliamentary Assembly. Resolution 2254 (2019)1. Media freedom as a condition for democratic elections, 2019a.

Council of Europe Parliamentary Assembly. Resolution 2300 (2019)1. Improving the protection of whistle-blowers all over Europe, 2019b.

Council of Europe, Consultative Committee of the Convention of the Protection of Individuals with Regard to Automatic Processing of Personal Data. Profiling and Convention 108+: Suggestions for an update, T-PD(2019)07BISrev, 2019.

Council of Europe, Directorate General of Democracy – European Committee on Democracy and Governance. The Compendium of the most relevant Council of Europe texts in the area of democracy, 2016.

Council of Europe, Directorate General of democracy and Political Affairs – Directorate of Democratic Institutions. Guidelines on transparency of e-enabled elections, 2011.

Council of Europe. Additional Protocol to the European Charter of Local Self-Government on the right to participate in the affairs of a local authority, 2009a.

_____. Algorithms and Human Rights. Study on the Human Rights Dimensions of Automated Data Processing Techniques and Possible Regulatory Implications, 2018. Available at https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html. Accessed: 5 may 2018.

_____. Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, 1997.

_____. Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, 2019a.

_____. Guidelines for civil participation in political decision making, CM(2017)83-final, 2017a.

_____. Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data (Convention 108+), 2018.Council of Europe.

_____. Recommendation CM/Rec(2001)10 on the European Code of Police Ethics, 2001.

_____. Recommendation CM/Rec(2004)15 on electronic governance ("e-governance"), 2004.

_____. Recommendation CM/Rec(2007)15 on measures concerning media coverage of election campaigns, 2007.

_____. Recommendation CM/Rec(2009)1 on electronic democracy (e-democracy), 2009b.

_____. Recommendation CM/Rec(2009)2 on the evaluation, auditing and monitoring of participation and participation policies at local and regional level, 2009c.

_____. Recommendation CM/Rec(2010)13 on the protection of individuals with regard to automatic processing of personal data in the context of profiling, 2010.

_____. Recommendation CM/Rec(2011)7 on a new notion of media, 2011.

_____. Recommendation CM/Rec(2014)7 on the protection of whistleblowers, 2014.

_____. Recommendation CM/Rec(2016)4 on the protection of journalism and safety of journalists and other media actors, 2016a.

_____. Recommendation CM/Rec(2016)5 on Internet freedom, 2016b.

_____. Recommendation CM/Rec(2017)5 on standards for e-voting, 2017b.

_____. Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership, 2018a.

_____. Recommendation CM/Rec(2018)4 on the participation of citizens in local public life, 2018b.

_____. Recommendation CM/Rec(2019)3 on supervision of local authorities' activities, 2019b.

_____. Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems, 2020b.

_____. Study on the Human Rights Dimensions of Automated Data Processing Techniques (in Particular Algorithms) and Possible Regulatory Implications, 2018c. Available at https://rm.coe.int/algorithms-and-humanrights-en-rev/16807956b5 Accessed: 15 jan. 2019.

Council of Europe. Graphical visualisation of the distribution of strategic and ethical frameworks relating to artificial intelligence, 2020a.

CRAWFORD K.; JOLER, V. Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources, 2018. Available at http://www.anatomyof.ai. Accessed: 27 dec. 2019.

CUMMINGS, M.L. et al. Chatham House Report. Artificial Intelligence and International Affairs Disruption Anticipated, 2018. Available at https://www.chathamhouse.org/sites/default/files/publications/research/2018-06-14-artificial--intelligence-international-affairs-cummings-roff-cukier-parakilas-bryce.pdf. Accessed: 21 mar. 2020.

DIAKOPOULPS. N. *Algorithmic Accountability Reporting: On the Investigation of Black Boxes*, 2013. Available at https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2. Accessed: 18 mar. 2018.

EDWARDS L.; VEALE, M. Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For. *Duke Law & Technology Review* 16, 18, 2017.

EU Code of Practice on Disinformation, 2018. Available at https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation. Accessed: 24 mar. 2021.

European Commission for Democracy trough Law (Venice Commission). Joint Report of the Venice Commission and of the Directorate of Information society and Actions Against Crime of the Directorate General of Human Rights and Rule of Law (DGI) on Digital Technologies and Elections, 2019.

European Commission, Networks, Content and Technology- Directorate-General for Communication. A Multi-Dimensional Approach to Disinformation Report of the Independent High-Level Group on Fake News and Online Disinformation, 2018.

European Commission. Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on Artificial Intelligence (Artificial

Intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, 2021.

European Union Agency for Fundamental Rights. #BigData: Discrimination in Data-Supported Decision Making, 2018.

European Union Agency for Fundamental Rights. Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights, 2019.

European Union Agency for Fundamental Rights. Preventing Unlawful Profiling Today and in the Future: A Guide, 2018.

EYKHOLT, K. et al. Robust Physical-World Attacks on Deep Learning Visual Classification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. Available at https://openaccess.thecvf.com/content_cvpr_2018/papers/Eykholt_Robust_Physical-World_Attacks_CVPR_2018_paper.pdf. Accessed: 23 apr. 2021.

FAYE JACOBSEN, A. The Right to Public Participation. A Human Rights Law Update. Issue Paper, 2013. Available at https://www.humanrights.dk/publications/right-public-participation-human-rights-law-update. Accessed: 14 jan. 2021.

FERRYMAN K.; Pitcan, M. *Fairness in Precision Medicine*, 2018. Available at https://datasociety.net/wp-content/uploads/2018/02/Data.Society.Fairness.In_.Precision.Medicine.Feb2018.FINAL-2.26.18.pdf. Accessed: 8 apr. 2018.

GANDY Jr., O.H.; Nemorin, S. Toward a Political Economy of Nudge: Smart City Variations. *Information, Communication & Society* 22, 2112, 2019.

GESLEY, J. Regulation of Artificial Intelligence in Selected Jurisdictions, 2019. Available at https://www.loc.gov/law/help/artificial-intelligence/index.php. Accessed: 30 dec. 2019.

GOODMAN E.; Powles, J. Urbanism Under Google: Lessons from Sidewalk Toronto. *Fordham Law Review* 88(2), 457, 2019.

GRABER, C.B. Artificial Intelligence, Affordances and Fundamental Rights. In Hildebrandt M.; O'Hara K. (eds) *Life and the Law in the Era of Data-Driven Agency*. Edward Elgar, 2020.

HAGENDORFF, T. The Ethics of AI Ethics: An Evaluation of Guidelines', *Minds and Machines* 30, 99, 2020.

HILDEBRANDT M.; Gutwirth, S. (eds) *Profiling the European Citizen: Cross-Disciplinary Perspectives*, Dordrecht, 2008.

HILDEBRANDT, M. Algorithmic Regulation and the Rule of Law. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 2018a.

_____. Primitives of Legal Protection in the Era of Data-Driven Platforms. *Georgetown Law Technology Review* 2, 252, 2018b.

_____ Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law* 20, 83, 2019.

_____. The Issue of Bias. The Framing Powers of Machine Learning. In Pelillo, M.; Scantamburlo, T. (eds) *Machines We Trust. Perspectives on Dependable AI* (MIT Press: Cambridge, MA) 2021.

Independent High-Level Group on Artificial Intelligence. Ethics Guidelines for Trustworthy AI, 2019. Available at https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthyai. Accessed: 2 mar. 2020.

JOBIN, A.; IENCA, M.; VAYENA, E. The Global Landscape of AI Ethics Guidelines', *Nature Machine Intelligence* 1, 389, 2019.

KAMINSKI M. E.; MALGIERI, G. Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. Association for Computing Machinery, 2020, doi: 10.1145/3351095.3372875.

KOLKMAN, D. The (in)Credibility of Algorithmic Models to Non-Experts. *Information, Communication & Society* 1, 2020, doi: 10.1080/1369118X.2020.1761860.

LOIDEAIN, N. N.; Adams, R. From Alexa to Siri and the GDPR: The Gendering of Virtual Personal Assistants and the Role of Data Protection Impact Assessments. *Computer Law & Security Review* 36, 2020, doi:10.1016/j.clsr.2019.105366.

LYNSKEY, O. Criminal Justice Profiling and EU Data Protection Law: Precarious Protection from Predictive Policing. *Int J Law Context*, 15, 162, 2019.

MAISLEY, N. The International Right of Rights? Article 25(a) of the ICCPR as a Human Right to Take Part in International Law-Making. *Eur. J. Int. Law* 28, 89, 2017.

MANHEIM, K.; KAPLAN, L. Artificial Intelligence: Risks to Privacy and Democracy. *Yale Journal of Law & Technology* 21, 106, 2019.

MANTELERO, A. AI and Big Data: A Blueprint for a Human Rights, Social and Ethical Impact Assessment. *Computer Law & Security Review* 34(4), 754, 2018.

_____. Artificial Intelligence and Data Protection: Challenges and Possible Remedies. Report on Artificial Intelligence. Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of personal data, 2019. Available at https://rm.coe.int/2018-lignes-directrices-sur--l-intelligence-artificielle-et-la-protecti/168098e1b7. Accessed: 20 feb. 2020.

_____. Personal Data for Decisional Purposes in the Age of Analytics: From an Individual to a Collective Dimension of Data Protection. *Computer Law & Security Review* 32(2), 238, 2016.

_____. Regulating AI Within the Human Rights Framework: A Roadmapping Methodology. In Czech et al. (eds) *European Yearbook on Human Rights 2020*, 477-502, 2020.

_____; Esposito. M.S. An Evidence-Based Methodology for Human Rights Impact Assessment (HRIA) in the Development of AI Data-Intensive Systems. *Computer Law & Sec. Rev.*, 41, 2021, DOI: 10.1016/j.clsr.2021.105561.

MANTELERO, A.; VACIAGO, G. Data Protection in a Big Data Society. Ideas for a Future Regulation. *Digital Investigation* 15, 104, 2015.

MEHR, H. Artificial Intelligence for Citizen Services and Government, 2017. Available at https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf. Accessed: 15 mar. 2021.

MEIJER A.; WESSELS M. Predictive Policing: Review of Benefits and Drawbacks. *Int J Publ Admin* 42, 1031, 2019.

MIKHAYLOV, J.; Esteve, M.; Campion, A. Artificial Intelligence for the Public Sector: Opportunities and Challenges of Cross-Sector Collaboration. *Phil. Trans. R. Soc. A*, 376, 2018, doi: 10.1098/rsta.2017.0357.

MITTELSTADT, B. From Individual to Group Privacy in Big Data Analytics. *Philosophy & Technology* 30, 475, 2017.

NEMITZ, P. Constitutional Democracy and Technology in the Age of Artificial Intelligence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378, 2018, doi: 10.1098/rsta.2018.0089.

NUNEZ, C. Artificial Intelligence and Legal Ethics: Whether AI Lawyers Can Make Ethical Decisions. *Tulane Journal of Technology and Intellectual Property* 20, 189, 2017.

OECD, Recommendation of the Council on Artificial Intelligence, 2019.

OSOBA, O.A.; Welser, W. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence, 2017. Available at https://www.rand.org/pubs/research_reports/RR1744.html. Accessed: 20 may 2020.

OSWALD, M. Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences'* 376, 2018, doi: 10.1098/rsta.2017.0359.

_____. Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 376, 2018, doi: 10.1098/rsta.2017.0359.

PASQUALE, F. *The Black Box Society. The Secret Algorithms That Control Money and Information*.Harvard University Press, 2015.

PASQUALE F.; CASHWELL, G. Prediction, Persuasion, and the Jurisprudence of Behaviourism. *University of Toronto Law Journal*, 68, 2018.

PERRY, W.L. et al. Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations, 2013. Available at https://www.rand.org/pubs/research_reports/RR233.html. Accessed: 30 mar. 2020.

Privacy International, Smart Cities: Utopian Vision, Dystopian Reality, 2017. Available at https://privacyinternational.org/sites/default/files/2017-12/Smart%20Cities-Utopian%20Vision%2C%20Dystopian%20Reality.pdf. Accessed: 12 may 2020.

RAAB, C.D. Information Privacy, Impact Assessment, and the Place of Ethics. *Computer Law & Security Review* 37, 2020, doi:10.1016/j.clsr.2020.105404.

RANCHORDÁS, S. Nudging Citizens through Technology in Smart Cities. *International Review of Law, Computers & Technology* 34(2), 254, 2020.

RE, R.M.; Solow-Niederman, A. Developing Artificially Intelligent Justice. *Stanford Technology Law Review* 22, 242, 2019.

RICHARDSON, R.; SCHULTZ, J.M.; CRAWFORD, K. Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. *New York University Law Review* 94, 42, 2019.

ROSENBAUM, D. P. The limits of hot spots policing. In Weisburd D.; Braga A.A. (eds) *Police innovation: contrasting perspectives*, Cambridge University Press, 2006.

SAVAGET, P.; CHIARINI, T.; Evans, S. Empowering Political Participation through Artificial Intelligence. *Science and Public Policy* 46, 369, 2019.

SCHRAG, Z. M. *Ethical Imperialism. Institutional Review Boards and the Social Sciences 1965-2009*. Baltimore, Johns Hopkins University Press, 2017.

SELBST, A. D. Disparate Impact in Big Data Policing. *Georgia Law Review* 52(1), 109, 2017.

SELBST, A. D.; BAROCAS, S. The Intuitive Appeal of Explainable Machines. *Fordham L. Rev.* 87, 1085, 2018.

SUNSTEIN, C. R. The Ethics of Nudging, *Yale Journal on Regulation* 32, 412, 2015.

_____. *Why Nudge? The Politics of Libertarian Paternalism*. Yale University Press, 2015.)

SUNSTEIN C.R.; THALER, R.H. Libertarian Paternalism in Not an Oxymoron', *University of Chicago Law Review* 70, 1159, 2003.

TAYLOR, L.; FLORIDI, L.; van der Sloot, B. (eds). *Group Privacy : New Challenges of Data Technologies*. Cham, Springer, 2017.

THALER, R. H.; SUNSTEIN, C. R. *Nudge. Improving Decisions about Health, Wealth, and Happiness.* New Haven, Yale University Press, 2008.

The United Nations (UN) Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe (OSCE) Representative on Freedom of the Media, the Organization of American States (OAS) Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples' Rights (ACHPR) Special Rapporteur on Freedom of Expression and Access to Information. Joint Declaration on "Fake News," Disinformation and Propaganda, 2017.

TUBARO, P.; CASILLI, A. A.; COVILLE, M. The Trainer, the Verifier, the Imitator: Three Ways in Which Human Platform Workers Support Artificial Intelligence. *Big Data & Society* 7(1), 2020, doi: 10.1177/2053951720919776.

UN Committee on Economic, Social and Cultural Rights (CESCR). General Comment No. 1: Reporting by States Parties, 27 July 1981.

UN Human Rights Committee (HRC). CCPR General Comment No. 25: The right to participate in public affairs, voting rights and the right of equal access to public service (Art. 25), CCPR/C/21/Rev.1/Add.7, 12 July 1996.

UNESCO. Draft Text of the Recommendation on the Ethics of Artificial Intelligence, 2021. Available at https://unesdoc.unesco.org/ark:/48223/pf0000377897. Accessed: 3 sep. 2021.

VAN BRAKEL R.; De Hert, P. Policing, Surveillance and Law in a Pre-Crime Society: Understanding the Consequences of Technology Based Strategies. *Cahiers Politiestudies, Jaargang* 3, 163, 2011.

VEALE, M.; BINNS, R. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society* 4(2), 2017, doi:10.1177/2053951717743530.

VERBEEK, P-P. Understanding and Designing the Morality of Things, Chicago-London, The University of Chicago Press, 2011.

VERONESE, A.; NUNES LOPES ESPIÑEIRA LEMOS, A. Trayectoria normativa de la inteligencia artificial en los países de Latinoamérica con un marco jurídico para la protección de datos: límites y posibilidades de las políticas integradoras. *Revista Latinoamericana de Economía y Sociedad Digital* 2, 2021. available at https://revistalatam.digital/article/210207/. Accessed: 27 aug. 2021.

WACHTER, S. Affinity Profiling and Discrimination by Association in Online Behavioural Advertising. *Berkeley Tech. L.J.*, 35(2), 367, 2021.

WEST, S.M.; WHITTAER, M.; Crawford, K. Discriminating Systems. Gender, Race, and Power in AI, 2019. Available at https://ainowinstitute.org/discriminatingsystems.pdf. Accessed: 15 may 2019.

ZALNIERIUTE, M.; Bennett Moses, L.; Williams, G. The Rule of Law and Automation of Government Decision-Making. *The Modern Law Review* 82(3), 425, 2019.

ZAVRŠNIK, A. Algorithmic Justice: Algorithms and Big Data in Criminal Justice Settings. *European Journal of Criminology* 1, 2019, doi:10.1177/1477370819876762.

ZUIDERVEEN BORGESIUS, F. Strengthening Legal Protection against Discrimination by Algorithms and Artificial Intelligence. *The International Journal of Human Rights* 24(10), 1572, 2020.

Sobre o autor:

Alessandro Mantelero | *E-mail:* alessandro.mantelero@polito.it

Associate Professor of Private Law and Law & Technology at the Polytechnic University of Turin, and Council of Europe Scientific Expert on AI, data protection and human rights. He is Associate Editor of Computer Law & Security Review and member of the Editorial Board of European Data Protection Law Review.

Artigo convidado.

Ethical Dimensions of the GDPR, AI Regulation, and Beyond[1]

HIELKE HIJMANS
Vrije Universiteit Brussels (Bélgica).

CHARLES RAAB[2]
University of Edinburgh (Reino Unido).

ABSTRACT: Our digital society is changing rapidly, with emerging new technologies such as artificial intelligence (AI) and machine learning, robotics, and the internet of things. These changes trigger new fundamental ethical questions[3] relating to privacy, data protection and other values, including human rights and the way they are affected by the extensive and intensive use of data for analytical and practical innovations. This article explores these ethical dimensions and the extent to which the European Union's General Data Protection Regulation[4] (GDPR) of 2018 takes ethics into account in relation to these socio-technical developments. More briefly, it looks similarly but more selectively at the EU's proposed AI Act of 2021, which aims to regulate AI in relation to levels of risk.[5] It concludes with some observations on desirable institutional arrangements for making and applying ethical judgements in the regulation of advanced technologies that use personal data.

1 An earlier version of this paper, "Ethical Dimensions of the GDPR" (2018) was prepared for inclusion in Cole, M. and Boehm, F. (eds.) (forthcoming) *Commentary on the General Data Protection Regulation*, Cheltenham: Edward Elgar. We are grateful to the editors for permission to use this material in a revised form.

2 Orcid: https://orcid.org/0000-0002-4579-0320.

3 Floridi calls this Information Ethics; the ethical impact of ICT on humans and society: Floridi, L. (2013), *The Ethics of Information*. Oxford: Oxford University Press.

4 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119/1.

5 Proposal from the European Commission for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative Acts, COM/2021/206 final.

1 INTRODUCTION

The GDPR gives effect to the fundamental right to data protection. It is a major landmark in the development of data protection, with a global reach far beyond the European Union. The extent to which, and the way in which, it embodies and promotes adherence to ethical norms and values beyond mere legal compliance sends important signals to those who process personal data in all domains of the public and private sectors. The GDPR also sets rules aiming at the protection of fundamental rights more generally[6], and is meant to be future-proof. An important perspective of the GDPR is that it recognises the contribution of technology to economic and social progress[7], but holds that technology should be developed in a responsible manner and, in particular, that individuals should have control over their personal data. Moreover, in recent years ethical issues have been increasingly recognised in the data protection community and in "data-driven" innovation and implementation[8]. The initiative of the European Data Protection Supervisor (EDPS), starting with an opinion on digital ethics and the establishment of an Ethics Advisory Group[9], is just one illustration of the wider "turn" to ethics; further examples will be mentioned later.

2 ETHICAL DIMENSIONS OF PRIVACY AND DATA PROTECTION

Ethical considerations are part of the GDPR, but they are sometimes hidden in its Articles. The GDPR has elements of principles-based regulation, and is not just rules-based[10]. We explore how the GDPR encourages ethical judgements, starting from the notion that the processing of personal data should be designed to serve mankind[11]. Ethical notions about good and bad behaviour lie behind legal stipulations of what must, can or cannot be done.

6 Recital 4.

7 *E.g.*, Recitals 2 and 6. This recognition is foregrounded even more in the AI Act proposal.

8 Some authors question the seriousness and sincerity of attempts to put ethical values into practice. See *e.g.*, references in Raab, C. (2020), "Information Privacy, Impact Assessment, and the Place of Ethics", *Computer Law and Security Review*, 37, July, https://authors.elsevier.com/sd/article/S0267-3649(20)30009-1.

9 See https://edps.europa.eu/data-protection/our-work/ethics_en; European Data Protection Supervisor, Opinion 4/2015 of 11 September 2015, "Towards a new digital ethics"; Press Release EDPS of 28 January 2016, https://edps.europa.eu/data-protection/our-work/subjects/ethics_en.

10 See Black, J. (2008) "Forms and Paradoxes of Principles Based Regulation", LSE Legal Studies Working Paper No.13/2008; Raab, C. (2012) "Regulating Surveillance: The Importance of Principles", p. 377-385 in Ball, K., Haggerty, K. and Lyon, D. (eds.), *Routledge Handbook of Surveillance Studies*, London: Routledge; Raab, C. (2017) "Information Privacy: Ethics and Accountability", p. 335-347 in Brand, C., Heesen, J., Kröber, B., Müller, U. and Potthast, T. (eds.), *Ethik in den Kulturen – Kulturen in der Ethik: Eine Festschrift für Regina Ammicht Quinn*, Tübingen: Narr Francke Attempto.

11 Recital 4.

Data protection laws are no different in this respect, in that they are based on ethical notions that underpin the fundamental rights of privacy and data protection. What the GDPR – as well as earlier generations of data protection laws – lays down in legislation incorporates ethical principles, although their prominence, explicitness and clarity is intermittent and fragmentary across the various Recitals and Articles. Moreover, the range of ethical principles and human rights upon which the GDPR could have drawn is wider.

Ethical values are often expressed in vague language and slogans that have only moral force, yet they may overlap with sometimes enforceable rights. The Charter of Fundamental Rights of the European Union[12] ("Charter") establishes (an undefined) "human dignity" as a bedrock in Article 1, saying: "Human dignity is inviolable. It must be respected and protected". Dignity is also emphasised with regard to the elderly (Article 25) and workers (Article 31). Dignity is also a key value underlying privacy and data protection. For the EDPS, human dignity was the main driver for dealing explicitly with ethics: it envisaged inserting this ethical claim into the "ecosystem" of data protection[13]. However, "dignity" is only mentioned once in the GDPR – in Article 88(2), concerning processing in the context of employment. Elsewhere in the GDPR, Recital 4 contains a very general ethical reference, saying: "The processing of personal data should be designed to serve mankind". On the other hand, the GDPR's emphasis on respect for the "fundamental rights and freedoms" of natural persons is ubiquitous, and the GDPR specifies the fundamental right of data protection in a number of detailed rights of the data subject that have legal effect[14]. The notion of fairness plays a key role in this respect. Fairness is an ethical dimension that is central to legal requirements for data protection under international and EU law, as well as national law. Article 8 Charter requires fair processing, whilst Article 5(1)(a)) GDPR elaborates this and associates it with transparency, although it can be debated whether transparency is an element of fairness or a separate requirement. In any case, "fairness" is a highly complex and ambiguous concept[15] and its relationship to other principles can be seen in various ways. We will touch upon fairness again later on.

12 Official Journal of the European Union, C/303/1.

13 European Data Protection Supervisor, Opinion 4/2015 of 11 September 2015, *Towards a new digital ethics: Data, dignity and technology*; EDPS Ethics Advisory Group Report, *Towards a Digital Ethics*, 2018, available on edps.europa.eu.

14 GDPR Chapter III.

15 Clifford, D. and Ausloos, J. (2018) "Data Protection and the Role of Fairness", *Yearbook of European Law,* 37,1: 130-187.

Another, more general, precept is respect for the rule of law, reflecting the ethical and legal roots of the EU and its Member States. Respect for the rule of law is reflected in data protection laws by providing that the processing of personal data must always rely on a legal ground (Article 8 Charter and Article 6 GDPR). The GDPR incorporates traditional data protection principles established in landmark documents, such as the Privacy Principles of the Organisation for Economic Co-operation and Development (OECD) of 1980[16] and the Council of Europe Convention No. 108 of 1981[17] that underlie Data Protection Directive 95/46 (the Directive) and national laws. These principles require that personal data should be fairly and lawfully collected for a valid purpose; accurate, relevant, and up-to date; not excessive in relation to its purpose; not retained for longer than needed for the purpose; collected with the knowledge and consent of the individual or otherwise on a legal basis; not communicated to third parties except under specified conditions that might include consent; kept under secure conditions; and accessible to the individual for amendment or challenge.

The GDPR goes further into the realm of ethics, emphasising principles that were not previously so prominent. These include transparency (or openness) and accountability, both of which address the ethical and practical relationship between controllers and processors of personal data and individuals, and reflect the OECD Principles. In another direction, the GDPR appears to go beyond the question of protecting individual's rights by drawing attention to the general *societal* interest in the protection of individuals" rights, for example in Article 57(1)(b), which says that a supervisory authority must "promote public awareness and understanding of the risks, rules, safeguards and rights in relation to processing". This codifies an existing practice in which supervisory authorities have for many years been engaged as societal "educators" in raising public awareness as part of their portfolio of tasks[18], although this role was not specified in the Directive.

16 Revised in 2013. See The OECD Privacy Framework http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf).

17 Council of Europe (1980), *Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data* (Convention 108). This was modernised in 2018 as Convention 108+, CM/Inf(2018)15-final. Note that Convention 108+ also embraces "dignity": its Explanatory Report states that: "Human dignity requires that safeguards be put in place when processing personal data, in order for individuals not to be treated as mere objects" (para 10.).

18 BENNETT, C. and RAAB, C. (2006) *The Governance of Privacy: Policy Instruments in Global Perspective*, 2nd edn. Cambridge, MA: The MIT Press, p. 139-40. Article 15 of Convention 108+ includes awareness-raising among the duties of supervisory authorities.

There are practical reasons for raising the level of public understanding about information processing and rights, but there is also an ethical dimension: awareness-raising could contribute to the shaping of societal conditions and an "ecosystem" for privacy and data protection that would be considered socially beneficial in the information age.

3 THE INCREASING PROMINENCE OF ETHICAL DISCOURSE

Beyond the GDPR and data protection more generally, in recent years there has been a proliferation of more directly ethical discourse in the data protection community among commercial and governmental data controllers, as well as supervisory authorities, featuring the formation of ethical codes of practice, ethics advisory processes and groups, and an increasing awareness of data ethics.[19] Alongside the EDPS initiative mentioned earlier, other examples include a report of the UK Information Commissioner's Office on big data, artificial intelligence, machine learning and data protection, with an emphasis on ethical approaches in data protection[20]. Further than this, and in the private sector, the Information Accountability Foundation[21] now includes ethics in its work among business and other data controllers[22]. However, it is often unclear what this ethical "turn" amounts to, how it relates to legal requirements, and what practices it enjoins on (or forbids to) whom; moreover, what its proponents mean by "ethics" is often not explained. Sceptics and critics of the efficacy of ethical and "responsible" approaches to information systems, data-driven innovations such as AI, and the like point to their powerlessness and the merely formulaic and pious terms in which ethical precepts are typically framed. These criticisms are often well argued, and the way in which data--related activities and regulatory or self-regulatory proposals by business or government are disguised by cosmetic ethical wrappers is easily shown. Nevertheless, doing what one should do with personal data, and refraining from what one should not do, are increasingly on the agenda and are likely to feature prominently in the implementation of the GDPR. Part of this agenda includes efforts for creating an effective ethical culture in which

19 See RAAB (2020), *supra* note 7, and the sources cited therein.

20 Information Commissioner"s Office (2017), *Big data, artificial intelligence, machine learning and data protection*, https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

21 Information Accountability Foundation (IAF) (2017), *Artificial Intelligence, Ethics and Enhanced Data Stewardship.*

22 RAAB (2017), *supra* note 9.

all stakeholders "do the right thing'[23], not as a substitute for enforceable legislation but as a complement to it.

Thinking ethically about the processing of personal data often causes dilemmas because judgment may be involved in situations where the line between "should" and "should not" is not clear. Moreover, ethics-based decision-making in "wicked" situations often points up the question of who is morally responsible for the "wrong" decision: a question that is not identical to the question of who is legally liable for it. Data controllers" declarations of the importance of ethical conduct in the "information age" may be a first step. There are also signs of genuine ferment and re-assessment as practitioners seek to understand the extra-legal implications of what they are doing, and to seek guidance towards decisions that involve something more than compliance with the law.

The search for ethical frameworks might denote the emergence of an information– (and information technology –) driven society in which risks and harms to human rights and social relations are taken more seriously. This development could also contest the prevailing approach in which convenience, commercial gain, and mission-oriented public policies take precedence in various contexts where personal data are processed. Of course, compliance with enforceable legal obligations should be most important; the strength of the GDPR lies in the fact that it contains clear rules and that breaches of the law are subject to sanctions, such as – but not only – considerable administrative fines..However, compliance needs reinforcement by ethical and rights-based precepts, if only to avoid a minimalist approach to data protection. Compliance with legal obligations should neither be driven by the prevailing approach mentioned above, nor by an attitude in which compliance would involve a mere check-list. Adopting ethical ways of data processing should be part of the DNA of organisations.

The role of ethics could become reinforced if the general public perceives that something is "not right', rather than 'not legal", in evaluating the data-driven processes to which they are increasingly involved. Recent controversies about the activities of Facebook, Instagram and other social media, as well as about Cambridge Analytica, serve to indicate that, more vocally than before, customers and citizens may press for ethically better

23 HIJMANS, H. (2018), "How to Enforce the GDPR in a Strategic, Consistent and Ethical Manner?", *European Data Protection Law Review* 1, with reference to the work of Christopher Hodges.

treatment rather than acquiescing in how they are dealt with by data controllers. However, perceiving and acting on the importance of ethical criteria is difficult in the contemporary global "datafied" society with its ubiquitous availability of personal information, especially when the means and purposes of data processing are not sufficiently clear[24]. In an information society in which power has shifted towards those who collect, share and hold large quantities of personal data[25], the GDPR might contribute significantly to giving these and other ethical principles some purchase to reduce the imbalance of power.

4 SURVEILLANCE: ITS ETHICAL IMPACT ON HUMANS

Surveillance is a key feature of a datafied society, and is facilitated or enabled by ubiquitous and often incomprehensible technology that is proving very difficult to regulate. There are many forms of surveillance or monitoring of individuals and groups: watching, listening, locating, detecting, and data monitoring[26] ('dataveillance", or the processing of personal data by business, governmental and social actors). Surveillance becomes even more intrusive when these forms are combined. The development of new technologies has expanded the possibilities and forms of surveillance, their interrelationships, and the ways in which they are used in a variety of contexts. These practices are driven by various economic, political, and social motivations that may be ethically dubious, and have a further negative effect on social values or rights, including privacy. The evolution of technologies brings forth increasing opportunities for surveillance that data protection laws and other safeguarding systems, including the GDPR, are designed to mitigate or sanction. However, new means of surveillance – coupled with, for example, automated decision-making on the basis of the data that is gathered – make many conventional safeguards obsolescent and irrelevant, stimulating further proposals for regulating AI.

---

24 MAYER-SCHÖNBERGER, V. and CUKIER, K. (2013) *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, [place of publication]: Eamon Dolan/Houghton Mifflin Harcourt.

25 HIJMANS, H. (2016) *The European Union as Guardian of Internet Privacy*, Law, Governance and Technology Series 31, at 3.6.

26 This section is adapted from RAAB, C. and WRIGHT, D. (2012) "Surveillance: Extending the Limits of Privacy Impact Assessment", p. 363-383 in Wright, D. and De Hert, P. (eds.), *Privacy Impact Assessment*. Dordrecht: Springer. For more detailed discussion, see Ball, K., Lyon D., Murakami Wood, D., Norris, C. and Raab, C. (2006) *A Report on the Surveillance Society*, for the Information Commissioner by the Surveillance Studies Network (SSN). http://ico.crl.uk.com/files/Surveillance%20society%20full%20report%20final.pdf; http://www.ico.gov.uk/news/current_topics.aspx; Monahan, T. (ed.) (2006). *Surveillance and Security: Technological Politics and Power in Everyday Life*. New York, NY: Routledge.

Surveillance and its regulation therefore require a closer look through the lenses of fresh and sometimes fundamental questions. Gary T. Marx calls for an ethics for the "new surveillance'[27]; he contends that ethical concerns – beyond procedural rules – have not tended to be raised about information collection or surveillance activities. He identifies three broad areas of importance: the means by which the data are collected; the contexts in which these processes take place; and the uses, purposes or goals for which the data are collected, or by which people are tracked, monitored, watched, overheard, or detected. Within these areas, Marx poses a number of ethical questions to be asked about surveillance systems. Some of the provocative – and political – questions he asks concern equality: "Is the means widely available or restricted to only the most wealthy, powerful, or technologically sophisticated?"; "Within a setting is the tactic broadly applied to all people or only to those less powerful or unable to resist?"; and "If there are means of resisting the provision of personal information are these means equally available, or restricted to the most privileged?". Another question relates to the essence of human dignity and harm to the individual: "Does the technique cross a personal boundary without permission (whether involving coercion or deception or a body, relational, or spatial border)?" Situational norms may be violated in ways that are not only unpleasant or uncomfortable, but that may harm individuals" interests, including their ability to either engage in certain social or political relationships or to withdraw from contacts by choice, as well as the exercise of their rights. In the wider context of machine learning, data and analytics may lead to beneficial or "pleasant" outcomes whilst also causing harm to the interests and rights of individuals.

Three causes of concern about surveillance are important in the debate on the ethical and legal dimensions of the GDPR:

a. *Legality*: the GDPR constitutes a system of rules and principles defining the legality of data processing. However, since these rules and principles cover a wide range of situations, they are by definition formulated in an open and general manner. Hence, the GDPR does not always provide certainty whether a particular surveillance practice is legal or not. For example, it does not specify the obligations of operators if surveillance is covert, as may be the case in some law-enforcement deployments, or if

27 MARX, G.T. (1998) "An Ethics for the New Surveillance", *The Information Society*, 14(3), p 171-86. See also Raab (2012), *supra* note 9.

surveillance takes place in a non-transparent manner, as may happen in commercial or public-service contexts. Moreover, surveillance in the public sector is to a large extent left to the discretion of national jurisdictions, under Article 23. Finally, a surveillance practice may be legally and ethically challenged because it is not compatible with the data protection principles of Article 5, such as fairness and transparency.

b. *Power implications*: surveillance implies a power relationship between the surveillant and the surveilled, in which the surveilled's power disadvantage may result in personal harm. This imbalance of power and its potential harmful consequences have important ethical dimensions. The GDPR requires in specific situations that these imbalances be taken into account, for instance in the balancing that is required when personal data are processed in the legitimate interest of the controller (Article 6(1)(f)), or in the application of the provisions on profiling and automated decision-making.[28]

c. *Differential effects of surveillance*: surveillance may adversely affect some individuals, groups, or categories of persons more than others. This may depend on how targets of surveillance are chosen (e.g., mass, or targeted on individuals or groups), or on how the placement of surveillance technologies inadvertently captures data from some people and not others because of the non-random way in which different persons, or kinds of person, may be exposed to the particular devices in question. Deliberate targeting, failure to anticipate differentials, mass data collection and analysis for public policies and purposes, and the maldistribution of privacy and other human rights caused by surveillance can all be questioned on ethical grounds, beyond what the law may or may not say.

Modern surveillance poses the question whether the GDPR and other legislative instruments can effectively keep surveillance within bounds that preserve human rights and liberties, or whether we need a renewed emphasis on the ethical principles that should underpin these legislative instruments, an emphasis that should also be inculcated in the training and

28 Notably, Article 22.

practice of technology innovators and the companies or governments in which they work. The results of such enquiry can inform the application of the GDPR or, in the longer term, a debate on the need for innovations in the GDPR, in other legislative instruments, or in non-legislative instruments such as codes of conduct, certification, or guidance by the European Data Protection Board (EDPB), national supervisory authorities (DPAs), and other official and unofficial bodies. The regulative effect of an elevated public awareness may also be enhanced by orienting it towards ethical and human--rights dimensions that have greater public resonance than do the legislated rules that are derived from them.

5 ETHICS AS PART OF DATA PROTECTION LAW

Section 2 above introduced the close relationship between data protection – and, more specifically the GDPR – and ethics. The fundamental right to data protection gives an individual a claim that her data is being processed in a fair manner[29]. Other obvious value notions within data protection are human dignity and personal autonomy[30]. In addition, ethical considerations play a role in the application of data protection law, including the GDPR. In a rapidly changing information society in which threats from the "new surveillance" are increasingly palpable, the application of the GDPR has to be based to a considerable extent on understandings of the way in which dignity, fairness and other values are implicated in the management processes whereby the law is put into practice in actual contexts, and on judgements about the compatibility of these situations with the realisation of, for example, a liberal democratic society based on the rule of law and human rights.

Ethical analysis is becoming part of the application of data protection law, situating the ethical content of data protection principles and the human-rights rationale of data protection law "on the ground". Some of the GDPR's Recitals imply ethical principles – for example, regarding discrimination and social disadvantage, in Recitals 71, 75 and 85 – and provide a handle for further analysis. Implementing the GDPR requires judgement; data protection law cannot – and should not – be merely technically applied. According to the Court of Justice of the EU (CJEU), this is even a key reason

29 Article 8(2) Charter. See also the three basis criteria of Article 5(1) GDPR: lawfulness, fairness and transparency.

30 See HIJMANS (2016), *supra* note 22: 2.8.2.

behind the independence of DPAs[31]. These authorities are required to "balance" individuals' rights with the economic interest of the free flow of data. Balancing, or reconciling, various interests is also at the core of the processing of personal data for the purposes of the legitimate interest pursued by the controller or a third party, in accordance with Article 6(1)(f). More generally, processing for "legitimate interest" purposes becomes increasingly important, requiring contextual assessments[32], including a balancing between fundamental rights and market dimensions. Balancing – however it is done, and despite the ambiguity of the term – should heed the ethical precept that data protection is a human right; its essence, and that of other rights and ethical values, should always be respected[33], especially where the "balance" is between a fundamental right and an economic or other interest, policy, or convenience[34].

We do not claim that the balancing of different interests always requires deep ethical judgements. However, balancing acquires an ethical dimension when it includes the weighing of moral values or human rights and, hence, a judgement about what is good and bad for an individual and society. This is specifically the case when the vague norms of the GDPR must be applied in situations of rapid change, in which the law itself does not give clear answers. Moreover, the GDPR itself contains a number of components that may require an ethical judgement, when applied. For example, Article 24, the general provision on the obligations of the controller, introduces a risk-based approach. It specifies that the controller must take into account "the risks of varying likelihood and severity for the rights and freedoms of natural persons". Recital 75 describes specific risks and harms to rights and freedoms of individuals that deserve protection. Its application may involve

31 CJEU, Case C-518/07, Commission v. Germany: 30.

32 NISSENBAUM, H. (2010) *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA: Stanford University Press.

33 Article 52(1) Charter.

34 For critical views of "balancing" where human rights and freedoms are involved, see Raab, C. (1999) "From Balancing to Steering: New Directions for Data Protection", p. 68-93 in Bennett, C. and Grant, R. (eds.), *Visions of Privacy: Policy Approaches for the Digital Age*, Toronto: University of Toronto Press; RAAB, C. (2017) "Security, Privacy and Oversight", p. 77-102 in Neal, A. (ed.), *Security in a Small Nation: Scotland, Democracy, Politics*. Cambridge: Open Book Publishers, https://www.openbookpublishers.com/resources/9781783742684/Security-Small-Nation-ch3.pdf; Chandler, J. (2009) "Privacy Versus National Security: Clarifying the Trade-Off", p. 121-138 in Kerr, I., STEEVES, V. and LUCOCK, C. (eds.), *Lessons from the Identity Trail: Anonymity, Privacy and Identity in a Networked Society*. Oxford: Oxford University Press; more generally, DWORKIN, R. (1977) *Taking Rights Seriously*, London: Duckworth; WALDRON, J. (2003) "Security and Liberty: The Image of Balance", *Journal of Political Philosophy*, 11(2): 191-210.

balancing these risks of data use with the benefits of data use in a developing information society, but this judgement is not straightforward.

In any balancing exercise, the beneficial contribution that privacy protection itself makes to society, and not only to the individual "data subject", has to be taken into account[35]. Understanding privacy protection's practical benefit to society is not always sufficiently appreciated in data protection law. For example, it could be argued that the quality of public services and official statistics is compromised if people have little confidence that their data would be adequately protected against breaches or unauthorised disclosure, and therefore give false information about themselves to doctors or public agencies. Similarly, the data subject's right to rectification under Article 16, underpinned by the "accuracy" requirement in Article 5(1)(d), illustrates the important contribution that the exercise of data subjects" rights – enshrined in data protection law – may make to the quality of data, whether held for commercial or public-service purposes: no--one would argue that defective databases are part of the common good. In addition, the contribution that privacy and human-rights protection make to the development of a "data-driven" economy and society by enhancing the public's trust and confidence in AI, data analytics and the use of algorithms is frequently proclaimed in programme rhetoric, although how far it is taken seriously is yet to be determined.

An ethical judgement to be made is whether one can require an individual to contribute to a "greater good" in another sense: more specifically, to hand over his or her personal data for increasing the quality of public services, for which the use of personal data using big data analytics may significantly benefit society by increasing the accuracy of services. This question frequently arises in the domain of health and medicine, with specific reference to the use of patient data in research. Examples can be found not only in health (how to prevent and cure cancer), but also in education (how to combat school drop-out) and in transport (how to meet the needs of passengers in public transport).[36] Ethically-loaded issues surrounding consent, the re-purposing of data, scientific research, and the transparency

35 See REGAN, P. (1995) *Legislating Privacy: Technology, Social Values, and Public Policy*, Chapel Hill, NC: The University of North Carolina Press, chapter 8; Raab, C. (2012), "Privacy, Social Values and the Public Interest", p. 129-151 in Busch, A. and Hofmann, J. (eds.) "Politik und die Regulierung von Information" ["Politics and the Regulation of Information"], *Politische Vierteljahresschrift* Sonderheft 46, Baden-Baden: Nomos Verlagsgesellschaft.

36 Examples from Information Commissioner"s Office (2017), *supra* note 19: 27.

of data-sharing are implicated in the pressures towards using personal data for "social benefit" or "public good". At the same time, individuals are entitled to control over their personal data and should be in a position to have a say, at least, in the disposition of their personal data. The claim of individuals is even stronger where there is a risk to their rights and freedoms, such as privacy, as explained in Recital 75 and as recognised in Article 9 on the processing of special categories of data, such as medical data.

6 ACCOUNTABILITY AND CORPORATE SOCIAL RESPONSIBILITY IN THE GDPR

Ethical behaviour is an integral element of accountability and of corporate social responsibility. A key concept is the principle of accountability as laid down in Article 24[37]. This requires data controllers to implement the necessary measures to ensure compliance and to be able to demonstrate that they have taken these measures. Both components are part of the legal responsibility of the controller, whereas the second component includes a procedural requirement to report, thereby demonstrating or giving an account of what they have done[38]. It is important to underline that accountability goes beyond legal responsibility. Legal responsibility concerns who is supposed to do what; it may involve legal implications for the performance or non-performance of roles and tasks, and for compliance with rules, as indicated in the GDPR. It may also involve authority relationships within organisational hierarchies, and demonstrating compliance to the supervisory authority and to the public at large[39], as far as this is explicitly required by the GDPR. Accountability, however, also implies that a responsible agent endeavours to respect the underlying principles of privacy and data protection and demonstrates its compliance or role-fulfilment even when this is not explicitly required by the GDPR.

Both the narrower legal responsibility and the wider principle of accountability involve ethical behaviour in two senses: in *performing*

37 Article 24 says "responsibility" instead of "accountability". The term "accountability" only appears in Article 5(2) and Recital 85. "Responsibility" is mainly used because there is no equivalent for "accountability" in a number of other languages (in particular, French). In the French language version, the term "responsabilité" is used for both responsibility and accountability. See the discussion of these two concepts in Raab, C. (2012) "The Meaning of "Accountability" in the Information Privacy Context", p. 15-32 and De Hert, P. (2012) "Accountability and System Responsibility: New Concepts in Data Protection Law and Human Rights Law", p. 193-232, both in Guagnin, D., HEMPEL, L., ILTEN, C., KROENER, I., NEYLAND, D. and POSTIGO, H. (eds.) *Managing Privacy through Accountability*, London: Palgrave Macmillan.

38 RAAB (2017), *supra* note 9.

39 For instance, by the annual report of the supervisory authority, as laid down in Article 59 GDPR.

according to the GDPR (insofar as its provisions are underpinned by ethical precepts), and – through giving an account of what one has done – in *explaining* this performance to others (i.e., the public or its representatives): in this way exemplifying one meaning of the principle of transparency. Accountability includes a risk-based approach[40] that may encourage data controllers to make ethical judgements, as they might do in undertaking a data protection impact assessment (DPIA): thus data controllers are required to evaluate risks to individuals" rights. Fairness is also recognised as an element of accountability, as in Article 5(1), where "fairly" is listed along with "lawfully and "in a transparent manner" with reference to the processing of personal data[41].

Moreover, accountability (including risk assessment) goes beyond compliance with the law. It is closely linked to corporate social responsibility, which makes companies responsible for their impact on society, integrating societal concerns into their business practices[42]. A parallel can be drawn with corporate social responsibility in the field of the environment. Others have developed the notion of enhanced accountability, also including ethical dimensions.[43]

There is also a link to responsible innovation (or responsible research and innovation). Where, as it is often claimed, the law cannot keep up with new technology, it may be a task for business to ensure responsible innovation.[44] To be responsible, innovation should be coupled with justified concerns for privacy, taking into account future uses of technologies and their possible privacy implications as well as consequences for other human rights[45]. Ethical organisations wish to do "the right thing'[46], and these good

40 As explained in earlier sections.

41 Centre for Information Policy Leadership (2015) "The Role of Enhanced Accountability in Creating a Sustainable Data-Driven Economy and Information Society", https://www.informationpolicycentre.com/cipl-white-papers.html; Raab (2017), *supra* note 9.

42 HIJMANS, H. (2016) *The European Union as Guardian of Internet Privacy*, Law, Governance and Technology Series 31: 6.14. See the remarks of the UK Information Commissioner, Elizabeth Denham, in her speech, "GDPR and Accountability", 17 January 2017, https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2017/01/gdpr-and-accountability/

43 Centre for Information Policy Leadership (2015), *supra* note 40.

44 JONES, M. (2017) "Does Technology Drive Law? The Dilemma of Technological Exceptionalism in Cyberlaw", https://ssrn.com/abstract=2981855.

45 Stahl, B. (2013) "Responsible research and innovation: The role of privacy in an emerging framework", *Science and Public Policy* 40: 708-716.

46 An essential element of ethical business practice; see Hodges, C. (2017 "Ethical Business Practice & Regulation: A Revolution in Regulation, Delivery, Enforcement and Compliance", *The 2017 Max Watson Lecture*, http://www.fljs.org.

intentions may give an ethical underpinning of activities relating to the processing of personal data. To quote Kant: "Nothing in the world – indeed nothing even beyond the world – can possibly be conceived which could be called good without qualification except a *good will*"[47]. Ethical behaviour also has a deontological dimension. The good will of organisations, as expressed in serious efforts, is part of accountability (or corporate social responsibility) under the GDPR. These principles are elaborated in various obligations for data controllers such as data protection by design and DPIA. It is safe to say that these specific instruments have an ethical dimension and should be applied to stimulate ethical behavior, not mainly as instruments for legal compliance. It is also possible, as some have shown, to go beyond a DPIA to develop an ethical or a social impact assessment, looking beyond privacy to the effects of surveillance on a wider range of human rights and values, both individual and collective[48].

An issue not yet discussed is whether this wide scope of accountability and responsibility – including ethical choices – implies that legal responsibility for compliance would also include moral responsibility for taking ethical considerations into account. This is an issue without an obvious solution. On the one hand, it is difficult to imagine that a large fine would be imposed by a DPA on a company for not taking ethical considerations into account beyond what is explicitly required by the GDPR. On the other hand, ethical choices may be the subject matter of a civil procedure on compensation and liability (as foreseen in Article 82) and could be included in judicial assessments in these cases.

7 HUMAN INTERVENTION AS A SPECIFIC ETHICAL COMPONENT OF THE GDPR

The ethical value of fairness is directly implicated in Article 22, which provides that individuals "shall have the right not to be subject to a decision based solely on automated processing", save for exceptions. The same Article provides that – where an exception to the main rule applies – an

47 KANT, I. (1959 [1785]) *Foundations of the Metaphysics of Morals*, trans. Beck, L. New York, NY: The Liberal Arts Press: 9, emphasis in original.

48 See, for example, WRIGHT, D. (2011) "A framework for the ethical impact assessment of information technology", *Ethics and Information Technology*, 13(3): 199-226; Wright, D. and Friedewald, M. (2013) "Integrating privacy and ethical impact assessment", *Science and Public Policy*, 40(6): 755-766; Wright, D. and Raab, C. (2012) "Constructing a Surveillance Impact Assessment", *Computer Law & Security Review*, 28(6): 613-626; Raab and Wright [2012], *supra* note 25.

individual has nevertheless a claim to obtain human intervention[49]. Article 22 reflects the view that important decisions for individuals should be made by humans, not by mathematical models, in what is sometimes called the "age of algorithms'[50]. Automated decision-making may dehumanise individuals or social processes[51]. Individuals must have the right to exercise influence over decision-making processes that significantly affect them[52]; whether they have the (equal) ability to exercise this influence – and whose responsibility it should be to enable them – raises further ethical issues[53]. As will be seen later on, topical developments such as AI and, related to this, machine learning, engage ethical questions that are reflected in the EU's proposed AI Act.

Machine learning can be defined as "the set of techniques and tools that allow computers to 'think' by creating mathematical algorithms based on accumulated data"[54]. To simplify it, the more data computers can use, the better they learn, insofar as the training data is relevant and unbiased. An argument in favour of automated decision-making is that it is supposed to be fairer or more efficient[55] or effective. A machine is said to be fairer because it eliminates human bias when decisions are made, for example, about whether to allow a person to enter a foreign country. It can also be more efficient: a well-trained algorithm will decrease the risks of false

49 In the post-Brexit UK, a government consultation paper (DCMS 2021) asked whether Article 22 should be removed from UK GDPR legislation, as had been proposed by the Prime Ministerially-appointed Taskforce on Innovation, Growth and Regulatory Reform (2021). The Information Commissioner"s Office (2021) expressed concern with this proposal, as it would weaken the protection of individuals, and proposed instead the extension of Article 22 to cover human review of partly automated systems.

50 This is a key narrative in O"Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, New York, NY: Crown Publishing Group/Penguin Random House.

51 Bygrave calls this dehumanisation of social processes. Bygrave, L. (2014), *Data Privacy Law, An International Perspective*, Oxford: Oxford University Press. See also Jones, M. (2017), "The right to a human in the loop: Political constructions of computer automation and personhood", *Social Studies of Science* 47(2): 216 –239.

52 See on this provision, MENDOZA, I. and BYGRAVE, L. (2017) "The Right not to be Subject to Automated Decisions based on Profiling", *University of Oslo Faculty of Law Legal Studies Research Paper Series* No. 2017-20, at 3.

53 Some of the questions Marx asked about surveillance relate to automated decision-making: *e.g.*, "Is there human review of machine-generated results?" and "Are people aware of the findings and how they were created?" See: Marx (1998), *supra* note 24. These have a very contemporary ring some 20 years after, especially concerning debates about the "right to an explanation" in the GDPR, *e.g.*, Wachter, S, Mittelstadt, B. and Floridi, L, (2017) "Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation", *International Data Privacy Law*, 7(2: 76-99.

54 LANDAU, D. (2016) "Artificial Intelligence and Machine Learning: How Computers Learn", *iQ*, 17 August, https://iq.intel.com/artificial-intelligence-and-machine-learning/

55 Fairness and efficiency are main themes in Zarsky, T. (2016) "The Trouble with Algorithmic Decisions: An Analytic Road Map to Examine Efficiency and Fairness", p. 118-132 in *Automated and Opaque Decision Making, Science, Technology, & Human Values*, 41(1), without concluding that machines are fairer or more efficient.

positives or false negatives. However, more importantly, algorithms change our perspective because of the absence of human scale. Transparency is rendered very difficult: it is not feasible to explain choices of the algorithm to the individual, and there are also doubts to what extent it is feasible to inform an individual in a meaningful manner about the logic involved[56]. A dilemma here relates to the extent to which individuals are entitled or able to challenge machine learning and decisions based on it for reasons of data protection (e.g., the requirement of fairness) and other relevant laws concerning, for example, equality, even if the learning is thought to be beneficial for society. It is very difficult for an individual to challenge the results produced through a machine in case of a biased or defective algorithm or because the individual's specific case does not fit the category in which the algorithm has classified her. Where an algorithm produces an unfair or incorrect outcome – in other words, a false positive or false negative – the burden of proof for the individual may be extremely high.

In certain cases, a solution was found, providing individual redress. Two examples are given here. The first involves the right to be forgotten or the right to be delisted, the issue at stake in three cases before the EU Court of Justice involving Google[57]. Where an algorithm produces certain results in a search engine leading to the disclosure of publications relating to an individual, the individual needs to act to undo these results if they prove to be unfair; in these cases, the disclosure of irrelevant and harmful information relating to an individual on the Internet, The CJEU required individual redress with human intervention. The second example is the case of Bettina Wulff, the former wife of the former President of Germany. When people searched her name on Google Search, terms like "prostitute" and "escort" appeared, because the information was published based on algorithms. She had to act to undo the negative effects, bringing a lawsuit against Google to have these results undone[58]. We also discover an ethical issue resulting from automated processing: should a provider of Internet services take (proactive) measures to avoid harm to individuals" privacy? A possible ethical answer could be

56 As required by Art 15(1)(h). See: Centre for Information Policy Leadership (2017), "Recommendations for Implementing Transparency, Consent and Legitimate Interest under the GDPR", 19 May, at 2.8, http://www.informationpolicycentre.com.

57 CJEU, Cases C-131/12, C-136/17 and C-507/17.

58 As explained in Jones (2017), *supra* note 50.

that we cannot rely on AI by itself; human involvement should remain in the lead, and the notion of "meaningful human control'[59] should prevail.

8 ETHICS IN THE EU'S PROPOSED AI ACT

Section 7 above touched on automatic decision-making, machine learning, algorithms and the ethical principles that are relevant to contemporary AI innovations. In all domains – *e.g.* health, education, law enforcement, transport, "smart" living, commerce, communications, finance, and many more – the advent of AI poses a strong challenge to the norms, values, laws and other regulatory instruments that were framed in terms of "data protection", "privacy" and personal data". In a society that is not only "datafied" but "data-driven", the adequacy of this conceptualisation and, more specifically, of the regulatory and governance ecosystem that evolved in terms of these frames come under question. The GDPR to a large extent renewed and strengthened the application of legal rules and the ethical precepts that relate to them, giving them wider geographical reach and reforming regulatory machinery, roles and relationships. The EU's proposal for AI regulation carries this ethical basis forward and presupposes as well as complements the GDPR's implementation, but focuses specifically on AI as the subject for a legal regime that is similarly informed by ethical norms. How adequately it fulfils this remit is open to debate.

Its context is important to note in assessing its merits and shortcomings. The proposal was not created in a vacuum, free of ethical discourse. It was closely preceded by the post-GDPR work of the European Commission's High-Level Expert Group on Artificial Intelligence (ECHLEG), which produced ethics guidelines for trustworthy AI[60]. These guidelines have become widely known, even amidst the welter of similar materials produced by many organisations in the flourishing "turn" to ethics, and are referred to in the draft AI Act's Explanatory Memorandum (sec. 3.2). ECHLEG identifies four principles, called "ethical imperatives" and alternatively referred to as rights, principles and values: respect for human autonomy, prevention of harm, fairness, and explicability. These principles are grounded in fundamental

59 JONES, M. (2020) "The Ironies of Automation Law: Tying Policy Knots with Fair Automation Practices Principles", *Vanderbilt Journal of Environment and Technology Law,* 18(1): 77-134.

60 European Commission High-Level Expert Group on Artificial Intelligence (2019) *Guidelines for Trustworthy AI*. For further discussion of this and other prominent sets of principles, see Raab (2020), *supra* note 7 and the sources cited therein.

rights, human-centricity, and concern for individual and societal benefit, and constitute an ethical top level of an elaborate architecture. This includes "requirements for trustworthy AI" under the headings of human agency and oversight, technical robustness and safety, privacy and data governance, transparency, diversity, non-discrimination and fairness, societal and environmental wellbeing, and accountability. Further down the chain there are more specific items, including technical and non-technical methods of operationalising trustworthy AI that involve 23 themes and a checklist of 60 questions and 69 sub-questions for an envisaged pilot exercise in analysing AI systems to test their ethical compliance.

The proposal is risk-based, with different types of AI categorised in terms of the risks they are deemed to pose. The proposal's heaviest regulation is carefully limited in scope to situations in which there is a "high risk" to fundamental rights and safety from any particular use of AI by itself for as a component of products, taking into consideration its "intended purpose" and not apparently covering unintended consequences. AI systems or products considered not to be of high risk are dealt with by voluntary measures such as codes of conduct (Recital 81 and Article 69), and there is also a category of prohibited AI (Article 5) that includes covert manipulations aimed at distorting a person's behaviour (but subject to a harm test), and certain law-enforcement deployments of remote real-time biometric identification systems that can, however, be authorised for use.[61]

In the eyes of some critics, the draft law is "stitched together from 1980s product safety regulation, fundamental rights protection, surveillance and consumer protection law'[62]. The risk, therefore, is that the fundamental rights and ethical values enshrined in the proposal's worthy intention may be compromised by the way in which the regulatory regime has been conceived and crafted. A powerful engine driving the proposal is the encouragement of the further development and implementation of AI in all domains without stifling innovation with onerous restrictions. As the Explanatory Memorandum (sec. 1.1) puts it, "this proposal presents a balanced and proportionate horizontal regulatory approach to AI that is limited to the minimum necessary requirements to address the risks and problems linked

61 For more detailed comment on how the wording of Article 5 provides loopholes, see Veale, M. and Zuiderveen Borgesius, F., (2021) "Demystifying the Draft EU Artificial Intelligence Act", *Computer Law Review International*, 22 (4). See also the review of their article by Kaminski, M. (2021), "The Law of AI", *Technology Law JOTWELL*, October 25.

62 Veale and Zuiderveen Borgesius (2021), *supra* note 60: 26.

to AI, without unduly constraining or hindering technological development or otherwise disproportionately increasing the cost of placing AI solutions on the market".

Adherence to ethical principles and the protection of fundamental rights is, to be sure, incorporated into the proposal, but it may be the weaker commitment in face of the powerful business and state interests in developing and deploying AI, and the application of the AI law may require more relevant machinery that is closer to the domain of rights protection. The mechanisms of the GDPR, protecting the rights of the data subject, could have served as an example for the application of practical ethics to regulating the broad field of AI in which the acknowledged potential risks and harms to individuals and groups would be addressed by solid regulatory and legal measures of broad applicability. Clarke's analysis indicates grounds for scepticism about the salience of the ethical notions in the draft AI Act. He develops a framework of ethical and practical requirements with 10 overarching "themes" and 50 "principles" and uses it as a yardstick for evaluating laws and policies that purport to promote or regulate ethical and responsible AI. [63] Whereas the ECHLEG guidelines score 74 % in conformity to his framework, the draft AI Act achieves 50%, with great variation across the fourfold risk levels for AI systems and many absent principles. It gains very low scores on many specific principles and relatively few high marks[64]. He shows that, across the four risk-level categories in the proposal, there are many gaps, loopholes, exceptions and other contributory factors that lead to an overall judgement that the proposed legislation is a failure in terms of principles and ethics[65].

It is worth noting that the EDPB and the European Data Protection Supervisor (EDPS) issued a Joint Opinion on the proposed AI Act, in which – while welcoming the proposal generally – they emphasise further

63 CLARKE, R. (2019) "Principles and Business Processes for Responsible AI", *Computer Law and Security Review*, 35(4): 410-422. The themes are: assess positive and negative impacts and implications; complement humans; ensure human control; ensure human safety and wellbeing; ensure consistency with human values and human rights; deliver transparency and auditability; embed quality assurance; exhibit robustness and resilience; ensure accountability for obligations; and enforce, and accept enforcement of liabilities and sanctions. The principles are mainly more granular specifications under each theme, combining mundane procedural steps (*e.g.*, conduct audits of safeguards and controls) with lofty ethical precepts (*e.g.*, respect for individual autonomy, freedom of choice, and right to self-determination).

64 Clarke, R. (2021) "The EC"s Proposal for Regulation of AI: Evaluation against a Consolidated Set of 50 Principles" Review Draft of 22 August 2021, http://www.rogerclarke.com/EC/AIP-EC21.html.

65 CLARKE (2021), *supra* note 63: 10.

apprehensions and call for the strengthening of safeguards[66]. The views of the EDPS are in particular weighty because the draft Act empowers it as the supervisory authority for the activities of the EU's institutions, agencies and bodies (Article 59(8) and as the market surveillance authority (Article 63(6)) for the same. The EDPB and EDPS are concerned particularly with data protection in the strict sense, including the relationship between an AI Act and the GDPR and the existing institutional data protection regime, but with wider AI implications as well. They write (para. 2): "Generating content, making predictions or taking a decision in an automated way, as AI systems do, by means of machine learning techniques or logic and probabilistic inference rules, is not the same as humans carrying out those activities, by means of creative or theoretical reasoning, bearing full responsibility for the consequences." Moreover, AI "will erode our capability to give a causal interpretation to outcomes, in such a way that the notions of transparency, human control, accountability and liability over results will be severely challenged" (para. 3).

With regard to ethics, and specifically discrimination, the Opinion supports bans on social scoring, the use of AI for the automated identification of individuals" biometric or behavioural features in publicly accessible spaces, and the use of AI to sort people into clusters based on personal characteristics. Article 21 of the Charter of Fundamental Rights of the European Union[67] is invoked to underpin non-discrimination, and other international rights-based documents are also cited more generally where the impact of AI on society and individual gives rise to apprehensions. The Opinion criticises the proposal's deficiency in addressing individuals" rights and remedies (para. 18), but they also go beyond the question of individuals" rights in pointing to wider societal and political risks: for example, "Post remote biometric identification in the context of a political protest is likely to have a significant chilling effect on the exercise of the fundamental rights and freedoms, such as freedom of assembly and association and more in general the founding principles of democracy" (para. 31). Concerning the important value of transparency, the Opinion supports – with suggested further strengthening – the proposal for establishing a public register of

66 EDPB-EDPS (2021), *Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act)*, 18 June, at https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021--proposal_en.

67 EDPB-EDPS (2021), *supra* note 65, Executive Summary and para. 5, and paras. 27-35.

stand-alone high-risk AI systems (paras. 69-72), although Clarke sees this instrument of transparency as more seriously deficient[68]. The Opinion also criticises the Article 52 exemption from transparency requirements of certain AI systems that are used in criminal law-enforcement, on grounds of sustaining the presumption of innocence as well as rights and freedoms through safeguards, checks and balances that are important in a "well-functioning democracy" (paras. 70-72).

Clarke looks closely at the ways in which the proposal's provisions fall short of implementing ethical and principled-based considerations[69], while Veale and Zuiderveen Borgesius also illustrate the draft's shortcomings in practice, given the wording of clauses that restrict the application of prohibitions or rules: for example, with regard to the conditions under which AI-related manipulation of persons may be outlawed, and with regard to the use of biometric identification systems[70]. Clarke further observes that the legislative proposal for AI does not even reflect the EU's own ECHLEG guidelines: "Key expressions… such as "Fairness", "Prevention of Harm", "Human Autonomy", "Human agency", "Explicability", "Explanation", "Well-Being" and "Auditability", are nowhere to be seen in the body of the Proposal'[71]. These are unfortunate deficiencies, but we may observe an even greater discrepancy between what is said in the document's own preliminary Explanatory Memorandum and Recitals, and what is laid down in the Articles of the proposed law itself. While "fairness" appears nowhere, "transparency", "trust", "trustworthiness" and many other ethical and rights-related values do frequently occur in the proposal's preliminary sections, but are not included in the Articles. The preliminaries are rhetorically impressive in their coverage of ethical principles and in their well-intentioned emphasis on safeguarding fundamental human rights concerning, for example, non-discrimination and privacy, which, it is acknowledged, may be harmed by AI applications in which surveillance and bias are prominent[72].. The devil is in the detail of the Articles, and the detail hides behind the ethical window-dressing. We emphasise that the enforcement of the AI Act will mainly take place on the basis of what is included in Articles, not on the basis of the intentions of the legislators.

68 CLARKE (2021), *supra* note 63: Appendix 1A.

69 CLARKE (2021), *supra* note 63: Appendices 1A and 1B.

70 Veale and Zuiderveen Borgesius (2021), *supra* note 60: 4-5, 7-9.

71 CLARKE (2021), *supra* note 63: 10.

72 See, for example, Recitals 33, 35-40.

9 IN CONCLUSION: WHO SHOULD TAKE THE LEAD IN MAKING ETHICAL JUDGEMENTS?

Making ethical judgements would require specifying criteria for judging "good", "bad", 'right", wrong", and other ethical evaluative categories, beyond the question of legal compliance. These are moral judgements in areas where there is not necessarily consensus in a society or across societies. These types of judgement also go beyond the scope of data protection. Nevertheless, ethical judgements are increasingly becoming an integral part of the application of the GDPR, as they are in many other contemporary pieces of legislation and policy developments in the fields of information technology and systems, AI, and "data science" more generally. As we have seen, these judgements are part of the accountability of data controllers wanting to do the right thing. Also, DPAs, when performing their wide range of duties under Article 57, should be guided by ethical considerations, which should also play a role in the sanctioning regime. To give an example, Article 83(2)(b) provides that the imposition of administrative fines shall, *e.g.*, be based on the intentional or negligent character of an infringement. This relates to the good will and ethical judgement of the controller.

The GDPR should be applied in a legitimate, effective and consistent manner[73]. Hence, ethical judgements should be made on the basis of ethical standards, but those that are comprised by the GDPR are not sufficiently specific, and the draft of the AI Act falls short of its promise. The dual objective of EU data protection (fundamental rights protection, but also free movement of data)[74] does not help either, and may simply point up the tensions between conflicting ethical principles, as well as their ambiguous meaning. This applies as well to the regulation of AI, in which there is considerable tension between the economic and political interests in technological innovation and exploitation in the one hand, and ethical values and human rights on the other. If ethical standards should, hence, be formulated and at least refine the general notions of the GDPR, the question is who should take the lead in this process.

The starting point could be that data protection is an area where expert bodies (the DPAs as supervisory authorities for data protection) with complete independence from the executive and other branches of government – and

73 HIJMANS, H. (2016) *The European Union as Guardian of Internet Privacy*, Law, Governance and Technology Series 31, 2016, *e.g.* at 10.3. The consistency mechanism is included in Chapter VII, Section 2 GDPR.

74 As explained by Lynskey, O. (2015) *The Foundations of EU Data Protection Law*, Oxford: Oxford University Press.

free from any external influence – play an important role in the application of the law. They are responsible for enforcement and also for giving guidance to stakeholders. Surveillance is an example where the protection of individuals should not depend on the political preferences of the day or – indeed – on majoritarian public opinion about the primacy of safety or public order[75]. This argument is even stronger in relation to ethical issues, and makes it sensible to consider organising standard-setting mechanisms located a certain distance from government. However, this does not mean that the DPAs are in an obvious position to fulfil this role. The EDPS took an important initiative by setting up a short-life Ethics Advisory Group[76], as noted earlier. This early step may be a recognition that – possibly – a supervisory authority for data protection is not fully equipped by itself to set standards for making ethical judgements. This article is not the place to develop a model for the governance of ethics within the context of the GDPR, nor indeed in other contexts that concern specific – but widely ramified – technologies and systems, such as AI. However, a few observations can be made.

First, although organisations themselves should play a significant part in making ethical judgements as part of their accountability[77] – a process that could not be replaced by the mechanical application of precisely formulated ethical standards – the development of more widely applicable ethical standards may prove to be useful, to promote legitimacy, effectiveness and consistency. As this is a domain where wide consensus on "good" and "bad" is needed, it would therefore be important to ensure that industry and civil society can play a role. As far as business may be in the lead, it should engage with other stakeholders (e.g., civil society, governments), as many private-sector organisations are now doing.

Second, ethics committees at regional (such as the EU), national or sectoral level could play a role in developing or even adopting ethical standards and in guiding ethical judgements in specific cases[78]. Since

75 For further reading, see HIJMANS, H. (2016) *The European Union as Guardian of Internet Privacy*, Law, Governance and Technology Series 31, Chapter 7.

76 See *supra* note 8.

77 The perspective of how organisations might address the application of ethical data processing to new technologies is at the core of the Information Accountability Foundation's paper by Abrams, M., Abrams, J., Cullen, P. and Goldstein, L. (2019), "Artificial Intelligence, Ethics and Enhanced Data Stewardship", *IEEE Security & Privacy*, 17(2).

78 An example is the European Group on Ethics in Science and New Technologies (set up by the European Commission) (2018), *Statement on Artificial Intelligence, Robotics and "Autonomous" Systems,* 9 March, https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1.

consistent approaches are needed in an information society with constant cross-border data flows, it is important that the big ethical questions should not be reserved for the national or even regional level. Where possible, consistency on ethical issues should not be limited to the EU itself and perceptions on ethics in other parts of the world should be included in the thinking process, but the possibility of worldwide, cross-cultural agreement on ethics and on the desirable extent to which they should prevail is likely to be remote.

Third, this should neither replace nor diminish the role of DPAs, which could develop views on how to include ethical considerations in the application of data protection laws, such as – within the European Union – the GDPR. An ethics committee (or multiple ethics committees) advising the DPAs could play a role here.

Sobre os autores:

Hielke Hijmans | *E-mail:* hielke.hijmans@gmail.com

President of the Litigation Chamber of the Belgian Data Protection Authority, Professor International and European Data Protection Law at the Vrije Universiteit Brussels.

Charles Raab | *E-mail:* c.d.raab@ed.ac.uk

Professorial Fellow, Politics and International Relations, School of Social and Political Science, University of Edinburgh.

Artigo convidado.

Vigilância e Relações de Poder – O Uso de Tecnologias de Reconhecimento Facial e Identificação Biométrica a Distância em Espaço Público e Impactos na Vida Pública

Surveillance and Power Relations – The Use of Facial Recognition Technologies and Remote Biometric Identification in Public Spaces and Impacts on Public Life

ANA CATARINA FONTES[1]
Technical University Munich – Alemanha.

CHRISTOPH LÜTGE[2]
Technical University Munich – Alemanha.

RESUMO: Os conceitos de espaço público e vida pública são abordados sob uma perspectiva espácio-antropológica com o objetivo de refletir sobre possíveis impactos da implementação de tecnologias de vigilância intrusivas. A vigilância massiva de espaços públicos impacta sobre a privacidade individual e as relações de poder estabelecidas, mas torna-se justificável sob o argumento de que garante a sua segurança e é utilizada para a manutenção da ordem pública. Em consequência, sistemas tecnológicos de crescente complexidade vêm sendo desenvolvidos e implementados em cidades por todo o mundo. A reflexão sobre impactos da implementação de sistemas de vigilância baseados em tecnologias de inteligência artificial (IA) inclui questões éticas relacionadas com valores culturais e direitos adquiridos, sendo questionado o uso de tecnologias de reconhecimento facial e identificação biométrica para a identificação de indivíduos em espaços publicamente acessíveis com base nos princípios de transparência, autonomia, proporcionalidade e equidade.

PALAVRAS-CHAVE: Reconhecimento facial; inteligência artificial (IA); vigilância; espaço público; vida pública.

ABSTRACT: The concepts of public space and public life are approached from a spatio-anthropological perspective and reflections are drawn on how they are impacted by the implementation of intrusive

1 Orcid: https://orcid.org/0000-0002-1341-8671.
2 Orcid: https://orcid.org/0000-0002-3870-4789.

surveillance technologies. The mass surveillance of public spaces undermines personal privacy and established power relations, but it becomes justifiable under the argument that it ensures the safety of such places and often used for law enforcement purposes. Therefore, increasingly complex and advanced surveillance systems based on technology, have been developed and deployed in cities around the globe. The reflection on impacts of the implementation of surveillance systems based on artificial intelligence (AI) technologies, includes addressing ethical issues related to cultural values and acquired rights and questioning the use of facial recognition and biometric identification technologies for the identification of individuals in publicly accessible spaces based on the principles of transparency, autonomy, proportionality and equity.

KEYWORDS: Face recognition; artificial intelligence (AI); surveillance; public space; public life.

INTRODUÇÃO

Na crítica ao modernismo, movimento que impulsionou um modelo de urbanismo centrado na especialização extrema do espaço urbano decorrente da segregação das funções na cidade, enceta-se um novo debate sobre o espaço público. O espaço público passa a ser visto como palco da vida social urbana, e o seu planejamento deve reger-se pela promoção de valores ligados à vida pública e a sociabilidades urbanas. Este debate, iniciado por autores como Jacobs (1961) e Sennett (1977), vem colocar o foco sobre parte humana da cidade e consignar o declínio da cidade ao declínio da vida pública. O espaço público é entendido como palco da vida pública e é produzido mediante representações e significados atribuídos socialmente com base em identidades e apropriações culturalmente enraizadas, tornando-se espelho da vida social urbana.

O espaço público é, assim, crucial para medir o pulso dos diversos atores e grupos que intervêm na construção de significados para a vida organizada em sociedade. A apreensão destes significados complexifica a circunscrição do espaço público a questões de propriedade, estratificando relações em que o espaço interseta a vida pública sob formas de apropriações particulares governadas contextualmente em função de aspetos culturais. O binômio espaço público/espaço privado é também incrementalmente desconstruído, quando aplicado em territórios incluídos que designam barreiras simbólicas, compreendendo formas de exercício de controle sobre o espaço e estratificando níveis de privacidade entre o público e o privado (Habraken, 1998). Na negociação do espaço sobre esses pressupostos, pro-

movem necessariamente situações de conflito (Mitchell, 1995) geridas entre as várias escalas dentro da cidade e atravessando as diferentes esferas espacial, política e social, em que a vida pública urbana se desenrola.

O Estado exerce o papel de regulador e representante último da ideia de público, acumulando funções como planeamento urbano, manutenção da ordem e segurança pública e proteção social de grupos vulneráveis. É, como tal, moderador e facilitador da vida pública, sendo a sua presença efetiva nas autoridades públicas, que, em última análise, supervisionam e exercem poder sobre o espaço por meio, por exemplo, da sua transformação material e da vigilância. Com efeito, vigiar é uma forma de exercer poder e estabelecer relações de controle.

Em Foucault (1979), evidencia-se que, mais do que vigiar, é a conjectura de estar a ser vigiado que incute um sentido de autodisciplina, condicionando o comportamento do indivíduo a corresponder à expectativa de quem o observa.

A conjunção destes pressupostos leva-nos a refletir sobre os impactos da implementação promovida por autoridades públicas de tecnologias baseadas em inteligência artificial (IA) para a identificação de indivíduos em espaços publicamente acessíveis. Com efeito, as potencialidades das tecnologias já desenvolvidas apresentam-se quase como ilimitadas na recolha e no processamento de dados sobre a vida quotidiana em espaço público, iterativa e cumulativamente capturando a identidade e traçando o perfil do indivíduo, capitalizando na possibilidade de "aceder à multidão" (de vigiar massas). Por outro lado, há de se refletir sobre os impactos que a implementação deste tipo de sistema pode ter sobre os significados e as expectativas projetados em relação ao espaço público e, por meio destes, como afeta as normas socialmente estabelecidas na autorregulação da vida pública. As questões éticas abordadas relacionam-se com valores culturais que encontram particular expressividade em espaços públicos e com a relação do Estado com os cidadãos, focando lógicas próprias da vida pública urbana "socialmente contratualizadas".

Para o caso em estudo, são analisados o princípio de transparência na relação entre Estado e cidadão, questões sobre a restrição de liberdades individuais e supressão de autonomia, que representam um risco na aplicação da lei internacional, nomeadamente para os direitos consagrados na Declaração Universal dos Direitos Humanos (United Nations, 1948). Torna-se ainda pertinente equacionar questões sobre proporcionalidade na avaliação

de custos e benefícios para a sociedade como um todo considerando riscos para o bem-estar societal.

O artigo fundamenta-se na literatura de referência para definir conceitos centrais como espaço público e vida pública, sob uma perspectiva espácio-antropológica, e contribui para o debate sobre questões éticas relacionadas com a implementação de tecnologias baseadas em IA, aportando reflexões sobre o caso do uso de tecnologias de reconhecimento facial e identificação biométrica por autoridades públicas em espaços públicos e publicamente acessíveis.

ESPAÇO PÚBLICO E VIDA PÚBLICA

A disseminação da ideia de espaços públicos e o alargamento da rede de espaços que correspondem à via pública da cidade desenvolvem-se a par dos valores e das transformações sociopolíticas e econômicas que levam ao estabelecimento das democracias ocidentais. Com efeito, a democratização do espaço público, que passa a ser o palco da vida pública para todos os cidadãos, efetiva-se com a abertura da economia e a abolição de barreiras geradas por classes sociais como forma de abrir o acesso à vida pública (Sennett, 1977). Depois da Segunda Guerra Mundial, o investimento do Estado na economia impactou nomeadamente sobre questões de planejamento urbano. A transformação de espaços urbanos proliferou por toda a Europa por meio de projetos de requalificação e renovação de espaços públicos.

No entanto, por volta dos anos 60-70, é retratado um novo ciclo de declínio do espaço público e seus significados para a vida pública (Jacobs, 1961; Sennett, 1977). Com a generalização do uso do automóvel e o desenvolvimento de novos sistemas de transporte, a expansão das cidades tende a consolidar o individualismo e a reger-se pela segregação funcional. As transformações estruturais testemunhadas são ainda de cariz social. Simmel (1973) refere como o anonimato e a alienação passam a definir formas de agir na vida pública associada a espaços públicos.

Arendt (1958) e Habermas (1978) referem-se a uma sociedade de massas (*mass society*) contemporizando uma homogeneidade aparente, radicada na padronização das interações sociais e na alienação política: enquanto este pressuposto poderia levar a confundir homogeneidade com coesão, a definição adquire uma conotação claramente negativa ao manifestar que, na diversidade, encontramos, afinal, maior partilha, com a coconstrução de significados para o espaço e vida pública. A diversidade é efetiva com a

afirmação de identidades num quadro não consensual, levando à ideia de espaço público como lugar de conflito (Mitchell, 1995).

Ao longo dos tempos, o espaço público vem acumulando significados como palco para a construção de identidades coletivas e manifestação de valores culturais, tornando-se produto das representações e práticas sociais a que lhe estão associadas e que ultrapassam o componente espacial, focando-se nas relações entre atores que interagem no e atuam sobre o espaço. É, portanto, um espaço produzido por meio de discursos, intenções, interpretações, apropriações, em que a definição passa mais pela forma como é socialmente construído do que por definições espaciais com base em barreiras arquitetónicas e de propriedade (Lefebvre, 1991).

Outros autores apresentam a impossibilidade de definir espaço público sem apresentá-lo num âmbito bidimensional, isto é, como oposto de espaço privado. No entanto, essa construção em binômio não é considerada estanque, esbatendo-se, ao invés, em ambiguidades e temporalidades com base na construção social e dinâmica de eventuais limites e barreiras (Habraken, 1998; Madanipour, *Public and private spaces of the city*, 2003).

O espaço privado encontra-se, em contexto urbano, predominantemente associado ao interior dos edifícios. A construção, a materialidade, traduz-se por definição na separação entre dois domínios, que prezam diferentes valores e ocupam dimensões complementares e tangentes. No espaço público, a presença e a possibilidade de encontros é admitida como casual, enquanto o espaço privado fica subordinado à questão da propriedade, um espaço de elevado escrutínio, onde a privacidade é um valor condescendido como direito inquestionável (Fontes, 2020).

Considerando que as duas esferas ou domínios são construções sociais, simetricamente espelham e são espelho da relação do indivíduo com o outro, e ainda que essa perspectiva seja altamente condicionada por fatores culturais, o contexto urbano parece oferecer ao espaço público condições particulares para a sua definição, desencadeando tipos de relações estritamente vinculadas aos urbanistas (Simmel, 1973; Hannerz, 1980; Benjamin, 1991). Hannerz (1980) refere como estes novos tipos de relações estritamente vinculadas a vivências urbanas dependem da oportunidade inadvertida de encontro, pois, sendo situacionais e casuais, relacionam-se com a dimensão da vida quotidiana que ocorre e é promovida em espaço público.

A consolidação do espaço público como espaço de sociabilidades urbanas gera-se mediante uma atratividade incrementada destes espaços, que funcionam como palcos da vida públicos. Nomeadamente, por meio da oferta diversificada de atividades e acontecimentos, em que se antecipa a possibilidade de encontro, estabelecendo redes de interação graduadas em diversos níveis de partilha e proximidade entre indivíduos e atores. Apesar de se tratar da esfera pública da vida em sociedade, é o indivíduo quem dita os termos, definindo regras e limites para integrar a rede de interações, nomeadamente com base em expectativas individuais sobre questões de privacidade.

A expectativa de permanecer anônimo na multidão é um dos mecanismos protetores da privacidade em espaço público. Na personagem do *flâneur*, Baudelaire (1991) ilustra como a cidade e a vida pública se tornam uma espécie de vitrine da modernidade. A vida pública é construída a partir da possibilidade de encontro com o outro, não obstante a passividade inerente como forma de interação. As relações sociais em espaço público podem resumir-se a acontecimentos casuais, sendo, no entanto, marcadamente relevantes para a produção do espaço público, enquanto simbólico e representacional para a definição da vida pública.

A dimensão do espaço público como espaço de encontro, de copresença ou de contato com o outro é defendida por Jacobs (1961) referindo-se a uma vida pública informal. O espaço público é palco da vida pública porque funciona como magneto agregador. Determinadas práticas relacionadas com consumo e o ócio encontram-se nele, particularmente promovidas ou mesmo exacerbadas. Por outro lado, a rede de estranhos anônimos que partilham, até certo ponto, estilos de vida e intenções sobre o espaço toma parte na própria paisagem, adicionando, com isso, uma camada de atratividade, ou seja, a copresença incrementa o valor do espaço público para a vida pública num ciclo constantemente reproduzido.

Outra dimensão que implica possibilidade de encontro e reunião em espaços publicamente acessíveis relaciona-se com acontecimentos mais ou menos organizados de expressão cultural e política. Nesta dimensão, o espaço público vem estabelecendo-se paulatinamente, enquanto promotor e consolidador de identidades coletivas. DaMatta (1997) apresenta o carnaval brasileiro como um desses momentos em que o espaço público é representação da vida pública ao fundir-se com um "mundo" cultural intangível. Agier (2011) apresenta uma proposta de classificação de situações com base

na relação que o indivíduo (citadino) estabelece com a cidade (nesse contexto, sinédoque de espaço público) e outros citadinos (sociabilidade).

Na proposta, o espaço público é considerado uma variável nesta equação para medir a intensidade das relações estabelecidas em determinadas tipologias de acontecimentos que desencadeiam ações coletivas. As diferentes tipologias propostas mediante as relações dominantes adquirem elevado significado para a definição de identidades coletivas e construção representacional e simbólica do espaço. As ruas da cidade tornam-se palco inicial de ações políticas e invenção cultural, por via da tomada de palavra pelas comunidades em manifestações. Esta apropriação do espaço público para afirmação social e cultural é, em si, uma forma de construir significados, representações e memórias coletivas. Os espaços marcados por esse tipo de acontecimento tornam-se, muitas vezes, icônicos para as comunidades, adquirindo significados que os distinguem de outras partes da cidade em associação a memórias coletivas, processos de emancipação social e política que projetam identidades urbanas vinculadas a determinados espaços urbanos. O espaço público como espaço de reunião, expressão e reivindicação tem vindo a ser determinante em movimentos de contestação política e questionamento de poderes instituídos, precisamente mediante a elevada visibilidade (vitrine da vida pública) e capacidade agregadora, onde cabem momentos de conflito e coesão social como duas facetas da negociação dinâmica e contínua do espaço.

O espaço público combina, assim, duas condições para a vida pública: é lugar de encontro e de expressão das relações e interações socialmente produzidas. Por um lado, estabelece a possibilidade de encontro numa rede que é alargada e representativa da própria sociedade e numa panóplia de situações que derivam da gestão dinâmica, contínua e mais ou menos informal, mas onde o indivíduo é sujeito soberano na determinação dos termos da interação. Por outro lado, consequentemente, potencializa-se a construção de significados coletivos numa espécie de coprodução espontânea e contínua do espaço. O espaço público é, assim, produzido socialmente e define-se por meio das interações situacionais e sociabilidades que acomoda numa estrutura continuadamente dinâmica.

RELAÇÕES DE PODER EM ESPAÇO PÚBLICO

A acessibilidade, forma como o acesso é gerido, é uma questão central na definição de espaço público. Um espaço é público quando é acessí-

vel, ou seja, quando o nível de escrutínio no seu acesso é tendencialmente nulo, por oposição ao espaço privado. Habraken (1998) refere-se a uma permeabilidade assimétrica entre espaço público e privado e percurso inverso (do privado para o público). Do espaço público para o privado, os níveis de escrutínio constroem-se em unidades que conformam territórios incluídos (onde o acesso é permitido a quem está neles admitido). No sentido oposto, do privado para o público, o acesso é progressivamente aberto a todos, enquanto o nível de escrutínio gradualmente desvanece. Por isso, quando falamos de espaço público, falamos de espaços publicamente acessíveis.

Neste contexto, podemos assumir que o espaço público serve à exaltação de valores democráticos, em representação das sociedades que neles se revêm, onde encontramos a diversidade, a tolerância e o respeito pela diferença. O espaço público define-se, então, como um espaço aberto e inclusivo. Esta vertente democrática de reivindicação de equidade no acesso e direito à *cidade* pode mesmo ultrapassar questões de propriedade, pondo em relação outros valores, como uso prolongado e significados construídos na proporção de direitos adquiridos (Mitchell, 1995).

Não obstante o pressuposto de que o espaço público é um espaço acessível e aberto, não é um espaço estável em termos de significados, pois, sendo socialmente produzido, fica sujeito às dinâmicas e às tensões sociais que o definem. A percepção do que é privado e do que é público depende ainda da perspectiva do indivíduo. O mesmo espaço pode ser simultaneamente privado para aqueles que não são nele admitidos e público para os que o entendem como um território incluído (Habraken, 1998), ou seja, os que são livres de entrar a qualquer momento. Nessa lógica, um espaço público poderia ser definido como um espaço utilizado por aqueles que não o controlam individualmente e onde o acesso é sempre garantido. Na polarização público/privado, qualquer espaço que seja exclusivo pode ser privado. No entanto, o nível de privacidade, entendido como o nível de contato com o público, pode assumir grandes variações. A distinção entre espaço privado e privacidade é um ponto importante para compreender relações de poder e controle sobre o espaço.

O espaço é fundamental no exercício do poder (Foucault, 1979); no entanto, o exercício de poder é também um fator para a construção de espaço, uma vez que é socialmente produzido. O conceito espaço público pode ainda se aplicar em dois sentidos: espaço do público, no sentido de aberto e acessível a todos, ou como espaço controlado e gerido pelo Estado,

enquanto representante último da sociedade como um todo (Madanipour, *Introduction*, 2010).

As relações de poder em espaço público processam-se a várias escalas de interação e ainda sob a forma de apropriações, que ocorrem numa base quotidiana. Estas põem em relação os diferentes atores que se definem por contraste, de acordo com papéis e intenções sobre o espaço, gerando situações de tensão, disputa e até conflito. As motivações evocam o direito de apropriação e acesso ao espaço público e por essa via à vida pública, sob determinados termos ou a necessidade de o gerir e regular. Com efeito, as relações de poder manifestam-se por meio das intenções de determinados grupos sobre o espaço, implicando negociação de controle sobre ele e a contestação de significados e representações. O Estado é um dos atores cujo principal papel é gerir e manter o espaço e a ordem pública. Vários outros grupos de atores identificam-se projetando narrativas que impactam diretamente na construção espácio-representacional.

Por meio do exercício de controle, o espaço torna-se alvo de disputa e espelho de contrastes e conflitos sociais. O exercício de controle passa, nomeadamente, pela imposição de narrativas e significados dominantes (por vezes incompatíveis com a aceitação de outros grupos, desencadeando processos de exclusão), pela transformação operativa do espaço no âmbito do planejamento e gestão urbana, ou até questões relacionadas com a manutenção da ordem pública (restrições ao trânsito, interdição de acesso por questões de segurança, entre outras).

A apropriação do espaço é também uma forma de exercer controle e reproduzir normas sociais para a vida pública. Apropriar implica reclamar (ainda que temporariamente) um espaço para uso pessoal. Quando um indivíduo se senta num banco num jardim público, mune-se de certo controle sobre esse mesmo espaço, pois passa a poder impor os termos em que outros indivíduos podem ter acesso a esse mesmo banco, seja por ele estar a ocupá-lo fisicamente, seja nos limites construídos em torno da sua expectativa de privacidade no momento em que efetua essa interação.

Com efeito, neste contexto, privacidade traduz-se na ação de excluir, restringir o contato, criar uma barreira protetora sobre o espaço pessoal, interditando-o ao poder do outro. Por oposição, invasão da privacidade significa (num sentido mais alargado) privar o outro da possibilidade de exercer controle sobre um espaço. A privacidade é, então, uma forma de expressar individualidade por meio da construção das barreiras necessárias para per-

mitir ao indivíduo tomar decisões, livre de forças coercivas (Squires, 1994), ou seja, é um processo individual de seleção sobre que informações que são partilháveis na esfera e na vida pública e as quais se mantêm privadas.

Os sem abrigo e os músicos de rua são dois grupos referidos na sua relação com o espaço público, mediante a ambivalente necessidade de reclamar um espaço para uso pessoal, mas estabelecendo contato com o outro, ditando termos específicos nessa interação. Quando nos referimos a grupos, há de se ressalvar que existem pelo menos duas formas de abordar a questão: uma perspectiva de classificação social, identificando indivíduos com características comuns, não significando que eles se sintam incluídos ou representados pela classificação estabelecida, ou uma iniciativa *bottom-up*, em que indivíduos se organizam em torno de interesses ou características comuns, podendo gerar movimentos em torno de causas. Nos últimos, existe uma coesão e estabilidade intrínsecas e conscientemente construídas que produzem fatores de segregação com a exclusão de outros indivíduos ou grupos de interesses incompatíveis. Este pode ser considerado um mecanismo para a construção de "espaços controlados" e escrutinados.

O sem-abrigo, reclamando o direito de aceder, apropriar-se e habitar o espaço público, exerce controle sobre este não somente por meio da sua presença, mas por meio da projeção de expectativas sobre privacidade que se elevam na proposição de utilizar espaços públicos para atividades, por norma, relacionadas com a esfera privada. Mitchell (1995) refere como as expectativas que diversos grupos projetam sobre o espaço se tornam incompatíveis, propondo a ideia de que o espaço público é inerentemente conflitual. No caso dos músicos de rua e também, por exemplo, dos taxistas, o contato com a vida pública é fundamental na construção da sua relação com o espaço (Jacobs, 1961; Habraken, 1998; Madanipour, *Public and private spaces of the city*, 2003). No entanto, os termos dessa interação não deixam de ser regulados por meio de normas sociais e marcação de limites na relação estabelecida, em função do respeito pela privacidade do indivíduo.

Se, por um lado, a presença em espaço público e a ideia de vida pública implicam a expectativa de encontro e partilha de espaço, por outro lado, os termos das interações em espaço público são construções sociais dinâmicas no tempo, baseando-se em limites consignados à privacidade do indivíduo e na negociação de intenções sobre o espaço. As relações de poder perpetuam-se na relação com o espaço numa constante dinâmica de contestação e reprodução. No processo, determinados atores ou grupos ob-

têm, em determinados momentos, maior ou menor poder de e para excluir e aceder ao espaço público.

VIGILÂNCIA E SEGURANÇA

Vigilância representa ambas as ideias: cuidar (*care*) e controlar (*control*) (Lyon, 2001). Vigilância é a forma de monitorizar as interações entre atores e indivíduos, particularmente relevante na manutenção da segurança e da ordem pública. Neste sentido, é um ponto central da vida pública, que ocorre em espaço público. Vimos, anteriormente, como o exercício de controle é inerente à apropriação do espaço; cumulativamente, a vigilância enquadra-se como uma das tipologias associadas ao exercício de controle, nas interações em espaço público.

A vigilância e a segurança (ou sensação de segurança) aparecem ainda vinculadas à necessidade de gerir e manter o espaço público. Jacobs (1961) explica como a fluidez na utilização da rua e dos edifícios fica potenciada por meio do comércio e da restauração, que atuam como facilitadores da interligação entre espaço público e privado. A diversidade de funções no espaço público amplia a oferta e a capacidade de atrair pessoas, considerando a sua copresença como autossuficiente para garantir uma vigilância natural, que resulta numa generalizada sensação de segurança. Este sistema de encontros, mantendo o anonimato da multidão, resulta numa rede de vigilância silenciosa e espontânea, protagonizada pelos que frequentam o espaço público e naturalmente o observam como efeito colateral. Neste modelo, os *vigilantes da rua* são os transeuntes que a percorrem, os consumidores que se servem dos espaços comerciais que lhe estão associados e os moradores que observam o espaço público, desde o refúgio da sua casa, espaço doméstico. A segurança da cidade está assim dependente da forma como o espaço público é frequentado, da densidade dos fluxos de pessoas que percorrem as ruas e consolidam a vida pública informal, permitindo a um visitante sentir-se incluído numa rede complexa e espontânea de vigilância.

Neste âmbito, não deixa de ser curioso notar como Lyon (2001, p. 15) relaciona o surgimento das sociedades de vigilância (*surveillance societies*) com o desaparecimento dos corpos (*disappearing bodies*). Entre os dois modelos dá-se uma transição de objeto vigiado: da silhueta humana – o anônimo na multidão, para informações sobre a vida quotidiana, ainda que fragmentárias, vinculadas a uma identidade. Entre eles, amplifica-se a possi-

bilidade de vigiar o outro para o poder de vigiar quem é o outro. Há ainda uma transferência da responsabilidade na ideia de vigiar como cuidar do outro, pois a relação construída informalmente por meio da copresença é substituída por sistemas de vigilância centralizados, como câmaras (CCTV) (Koskela, "The gaze without eyes": video-surveillance and the changing nature of urban space, 2000).

Foucault (1979) evoca que vigiar é apenas o primeiro passo na esfera de atuação do panóptico. Para exercer disciplina, um segundo passo fundamental é identificar o indivíduo e traçar o seu perfil/registro criminal a fim de atuar de acordo com informação acumulada ao longo do tempo. Com efeito, o poder sobre o outro advém da informação recolhida e acumulada e do controle permitido pela vigilância centralizada, mais ainda pela sensação construída intencionalmente de estar constantemente sob vigilância. Na forma como o sistema é desenhado, os mecanismos de vigilância estão constantemente expostos e indiciam uma espécie de omnipresença de quem vigia, criando um constante estado de alerta radicado na ameaça de estar a ser vigiado e na impossibilidade de poder aferir a realidade. Koskela (2003) descreve a hipótese como uma situação de visibilidade bilateral, em que o indivíduo fica permanentemente exposto a ser observado, tal como os mecanismos de controle. Os últimos são não só continuadamente perscrutáveis, mas intencionalmente visíveis.

Aqui vale introduzir o Estado como ator, que pode potencialmente desequilibrar as relações de poder por questões de soberania. A manutenção da segurança e da ordem pública entram na esfera de atuação das autoridades públicas, em última análise, responsáveis pela gestão de conflitos e controle da criminalidade. Em espaço público, a interferência do Estado, nomeadamente por meio de ações policiais, é justificada na patrulha e vigilância do espaço, sentidas em conformidade com a aplicação da lei.

O uso de câmaras de vigilância (CCTV) constituiu uma nova etapa, considerando os mecanismos disponíveis para vigiar espaços públicos com o objetivo de reduzir a criminalidade e, portanto, garantir a segurança. No entanto, vários estudos vêm apontando como estes sistemas de vigilância não são eficazes em todas as situações (Fyfe; Bannister, 1998; Ditton; Short, 2006) e levantam várias questões éticas na sua implementação, relacionadas não só com questões de privacidade, mas ainda com questões de proporcionalidade, transparência, discriminação (baseada em estereótipos e preconceitos) e exclusão (Taylor, 2002; Kostela, *"Cam Era" – The contemporary urban panopticon*, 2003; Norris; Armstrong, 2006).

AS POTENCIALIDADES DA INTELIGÊNCIA ARTIFICIAL

Uma das limitações dos sistemas CCTV é a dependência da constante intervenção humana na observação e interpretação das imagens captadas (Ditton; Short, 2006). Com a integração de sistemas baseados em inteligência artificial (IA) promovida pelos avanços no campo da identificação biométrica e, em particular, nas tecnologias de reconhecimento facial, surgem novas potencialidades e aplicações para a vigilância de espaços. As tecnologias de reconhecimento facial são uma das possíveis formas de identificação biométrica, que se baseiam no reconhecimento de características físicas e/ou comportamentais únicas ao indivíduo para identificá-lo, ou seja, conhecer a sua identidade (Jinu; Sheeja, 2019).

Assim, a principal aplicação desta tecnologia, no âmbito dos sistemas de vigilância, é na identificação de indivíduos (inclusive em tempo real e a distância) por meio da comparação/correspondência entre informação recolhida e uma base de dados (de suspeitos procurados). A IA permite não apenas a automatização do processo, mas o seu constante aperfeiçoamento, pois o sistema "aprende" com os dados e processos realizados, melhorando a sua *performance* e precisão. Permite ainda combinar, no mesmo, sistema várias fontes de informação biométrica de forma a triangular dados e compensar lacunas (Si; Zhang; Li; Tan; Shao; Yang, 2020).

Outra limitação que a integração de tecnologia de IA em sistemas CCTV pode ajudar a ultrapassar é o (des)bloqueio da possibilidade de navegar e filtrar grandes quantidades de informação. Por meio da identificação do indivíduo, mas também da padronização de comportamentos considerados de interesse, o sistema viabiliza rapidamente a pesquisa de informação numa base de dados histórica, devolvendo os resultados pretendidos. Além da identificação de indivíduos, os sistemas de IA podem aprender a interpretar e prever ações humanas e a classificá-las como "normal", "anormal" ou "danosa/nociva" (Roo, 2011).

Os sistemas CCTV são sistemas passivos que dependem da intervenção humana para interpretar as imagens capturadas. Com a integração de IA, torna-se possível o processamento automático da informação, aumentando o potencial da tecnologia na resposta à quantidade de dados recolhidos e conferindo-lhe autonomia para tomar decisões e interpretar informação em substituição da ação humana.

QUESTÕES ÉTICAS, RISCOS E IMPACTOS

A implementação de sistemas de reconhecimento facial em espaços publicamente acessíveis tem sido acompanhada de reflexões sobre pertinência e justificabilidade. Estas referem-se tanto ao estado de desenvolvimento e fiabilidade da tecnologia (ou seja, está suficientemente avançada para apresentar resultados fidedignos?), como às implicações sociais e aos impactos sobre valores democráticos e direitos adquiridos, implícitos na relação do indivíduo com o espaço público e vida pública e com o aparelho estatal.

Muitos destes pressupostos não estão diretamente vinculados à tecnologia em si, mas têm, antes, raízes nos princípios que governam as estruturas sociopolíticas, em que o Estado se assume como supervisor e regulador da ordem pública. Em função da relação do Estado com os cidadãos, pode ser questionado até que ponto é justificável ou mesmo aceitável a vigilância massiva de espaços públicos para matérias de segurança e manutenção da ordem pública. Outro receio é o fato de a tecnologia funcionar por meio da recolha de informações pessoais sobre a vida quotidiana do indivíduo, que, além de impactarem a sua privacidade, munem o Estado e outras entidades em posse desses dados, de poder sobre o indivíduo, podendo comprometer a sua autonomia e acentuar assimetrias de poder, nomeadamente, desencadeando situações de discriminação, coerção e exclusão.

Com efeito, os argumentos políticos para a implementação de tecnologias de reconhecimento facial e identificação biométrica em espaços públicos têm vindo a esbarrar com movimentos de contestação e crítica social. Várias campanhas e relatórios[3], com particular ênfase no Reino Unido (de acordo com as iniciativas de implementação e repercussões nos media[4]),

3 Alguns exemplos de campanhas pelo banimento do uso de tecnologias de reconhecimento facial em espaços públicos: Reclaim your face, Stop facial recognition, Kameras Stoppen, Gesichtserkennung Stoppen. Relatórios sobre o tema: Ada Lovelace Institute (2019) Beyond face value: public attitudes to facial recognition technology. Disponível em: https://www.adalovelaceinstitute.org/wp-content/uploads/2019/09/Public-attitudes-to-facial-recognition-technology_v.FINAL_.pdfeinstitute.org; EDRi (2021) The rise and rise of biometric mass surveillance in the EU. Disponível em: https://edri.org/wp-content/uploads/2021/07/EDRI_RISE_REPORT.pdf; EU Agency for Fundamental Rights (2019) Facial recognition technology: fundamental rights considerations in the context of law enforcement. Disponível em: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2019-facial-recognition-technology-focus-paper-1_en.pdf – of law enforcement (europa.eu); FUSSEY, P.; MURRAY, D. (2019). Independent Report on the London Metropolitan Police Service's Trial of Live Facial Recognition Technology. University of Essex, Human Rights Centre. Disponível em: https://48ba3m4eh2bf2sksp43rq8kk-wpengine.netdna-ssl.com/wp-content/uploads/2019/07/London-Met-Police-Trial-of-Facial-Recognition-Tech-Report.pdf.

4 MORRIS, Steven. Office worker launches UK's first police facial recognition legal action. *The Guardian*, May 21, 2019. Disponível em: https://www.theguardian.com/technology/2019/may/21/office-worker-launches-

colocaram o tema na mira pública à qual organizações como a Comissão Europeia não ficaram indiferentes.

Por meio do Regulamento Inteligência Artificial (*Artificial Intelligence Act*) (COM, 2021), veio também a proposta de regulação do uso de tecnologias de identificação biométrica a distância, considerando a proibição da sua implementação em espaços publicamente acessíveis e em tempo real, apesar das exceções previstas (busca por vítima de crime, ameaça sobre a vida humana – integridade física ou terrorismo, mandatos de captura europeus). Em resposta à proposta, a Comissão de Proteção de Dados e Supervisor Europeu (*European Data Protection Board* e *European Data Protection Supervisor*) emitiu um parecer no qual considera a implementação de sistemas de identificação biométrica (a distância) de indivíduos em espaços publicamente acessíveis um elevado risco de intrusão na vida privada dos indivíduos, com impactos severos sobre a expectativa das populações de permanecer anônimos em espaço público. Recomendam, em consequência, o banimento geral de qualquer uso de IA para o reconhecimento automatizado de características humanas em espaços publicamente acessíveis e para a categorização ou classificação de indivíduos de acordo com proveniência étnica, gênero, orientações políticas ou sexuais, bem como qualquer outra forma de discriminação (EDPB EDPS, 2021, p. 2-3).

No sentido de abordar o tema desde uma perspectiva ética, focando riscos e impactos, apoiamo-nos nas diretrizes propostas em *Ethics Guidelines for Trustworthy AI* (COM, 2019) e em *AI4People – An Ethical*

uks-first-police-facial-recognition-legal-actionl recognition; DODD, Vikram. UK police use of facial recognition technology a failure, says report. *The Guardian*, May 15, 2019. Disponível em: https://www.theguardian.com/uk-news/2018/may/15/uk-police-use-of-facial-recognition-technology-failure; TAYLOR, Josh. Calls to stop NSW police trial of national facial recognition system over lack of legal safeguards. *The Guardian*, Jun. 30, 2021. Disponível em: https://www.theguardian.com/australia-news/2021/jul/01/calls-to-stop-nsw-police-trial-of-national-facial-recognition-system-over-lack-of-legal-safeguardsck of legal safeguards. Australian Police and Policing, *The Guardian*; GRIERSON, Jamie. Police trials of facial recognition backed by home secretary. *The Guardian*, Jul. 12, 2019. Disponível em: https://www.theguardian.com/uk-news/2019/jul/12/police-trials-facial-recognition-home-secretary-sajid-javid-technology-human-rightsrdian; SABBAGH, Dan. South Wales police lose landmark facial recognition case. *The Guardian*, Aug. 11, 2020. Disponível em: https://www.theguardian.com/technology/2020/aug/11/south-wales-police-lose-landmark-facial-recognition-case recognition; TAYLOR, Josh. Major breach found in biometrics system used by banks, UK police and defence firms. *The Guardian*, Aug. 14, 2019. Disponível em: https://www.theguardian.com/technology/2019/aug/14/major-breach-found-in-biometrics-system-used-by-banks-uk-police-and-defence-firmsecurity; SABBAGH, Dan. Facial recognition technology scrapped at King's Cross site. *The Guardian*, Sep. 2, 2019. Disponível em: https://www.theguardian.com/technology/2019/sep/02/facial-recognition-technology-scrapped-at-kings-cross-developmentte. Facial recognition. *The Guardian*; BOWCOTT, Owen. Police use of facial recognition is legal, Cardiff high court rules. *The Guardian*, Sep. 4, 2019. Disponível em: https://www.theguardian.com/technology/2019/sep/04/police-use-of-facial-recognition-is-legal-cardiff-high-court-rulesition; DAVIS, David. Facial recognition technology threatens to end all individual privacy. *The Guardian*, Sep. 20, 2019. Disponível em: https://www.theguardian.com/commentisfree/2019/sep/20/facial-recognition-technology-privacy.

Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations (Floridi *et al.*, 2018).

TRANSPARÊNCIA E AUTONOMIA

De acordo com Florini (2007), transparência refere-se ao grau no qual a informação se encontra disponível para os que estão fora do sistema, capacitando-os da possibilidade de ter vozes informadas nas decisões e/ou avaliar decisões tomadas por quem "pertence ao sistema". Essa definição geral de transparência serve-nos para avaliar como a implementação de sistemas de vigilância intrusivos, tais como sistemas que permitem identificar indivíduos a distância e em tempo real, em espaço público, é conduzida por autoridades públicas.

Koskela (2003) refere que uma das condições para o sucesso do panóptico é exatamente a impossibilidade de o indivíduo verificar o sistema de vigilância, isto é, saber quando e por quem está a ser observado, gerando nele um constante estado de alerta. Ainda que o funcionamento próprio das forças policiais seja em determinadas situações pelo secretismo (no sentido de surpreender o suspeito), quando se trata da vigilância massiva de espaços públicos, a opacidade sobre onde, quando e com que finalidade um sistema é implementado representa riscos agravados para a autonomia do indivíduo.

O princípio do respeito pela autonomia humana baseia-se na concessão da autodeterminação plena e efetiva sobre si próprio e na participação no processo democrático. Os sistemas de IA não devem subordinar, coagir, enganar, manipular, condicionar ou arregimentar injustificadamente os seres humanos. Em vez disso, os sistemas de IA devem ser concebidos para aumentar, complementar e capacitar as competências cognitivas, sociais e culturais dos seres humanas (COM, 2019).

Privar o indivíduo de ser informado de que a sua identidade é rastreada em circuitos que se confinam ao espaço público é privá-lo da capacidade de tomar decisões autônomas sobre a forma como conduz atividades da sua vida quotidiana e da sua vida pública. Essa situação assume proporções amplificadas, recuperando as ideias apresentadas sobre a importância do espaço público na organização da vida pública e na consolidação de identidades coletivas e de valores democráticos. A implementação ética de sistemas de vigilância implica assegurar transparência na relação entre o Estado e o indivíduo.

Na interação com sistemas de IA, os humanos devem sempre reter o poder e a liberdade para tomar decisões informadas (Floridi *et al.*, 2018). Esse cenário é válido tanto para quem implementa e opera a tecnologia como para quem é afetado por ela.

A tecnologia, em si, deve ser desenvolvida com base no princípio da explicabilidade. Explicabilidade introduz um ângulo adicional ao abordar questões de transparência desde a perspectiva do processo de desenvolvimento da tecnologia com base em IA. A complexidade e a autonomia inerentes à IA tendem a tornar os processos e as decisões automatizadas opacos, arriscando ser totalmente ininteligíveis para o humano. Uma vez mais, desde essa perspectiva, o foco está na descodificação de todas as informações necessárias para que, no processo de interação, o humano esteja capacitado a tomar decisões informadas e em posição de supervisionar e agir de forma crítica sobre o funcionamento da tecnologia, assegurando que a sua autonomia é respeitada.

No fomento da transparência na implementação desse tipo de sistema, o paradoxo transparência-eficiência estabelece-se com um possível desafio. A vigilância é mais eficiente quando é invisível ou quando se torna omnipresente, atuando por excesso. Ambos os extremos representam riscos sobre questões de transparência e para a autonomia humana, sendo, neste paradoxo, difícil gerir questões de proporcionalidade que abordaremos mais à frente.

Quando o sistema é implementado pontualmente e sendo facultada essa informação, vários autores têm verificado que, em vez de representar uma solução para a segurança do espaço público, apenas dispersam atividades ligadas à criminalidade para outros pontos não vigiados da cidade (Ditton; Short, 2006).

Na ideia de rede de vigilância informal introduzida por Jacobs (1961), existe um equilíbrio na distribuição de poder sobre o espaço: em primeiro lugar, porque não se rege pela intenção explícita de vigiar o outro; em segundo lugar, porque todos os indivíduos interagem em circunstâncias semelhantes. Quando a vigilância é centralizada, o Estado está em posição de controle permanente sobre o espaço público. Neste caso, a vigilância vem acentuar assimetrias de poder, pois quem é observado não está em igualdade de circunstâncias em relação a quem observa.

Como base nos argumentos expostos, podemos assumir que vigilância massiva de espaços públicos representa um risco para o bem-estar socie-

tal (sociedade e democracia). Enquanto quem vigia se torna uma entidade progressivamente mais distribuída e opaca, quem é vigiado torna-se mais visível e previsível, estudado e perscrutado de forma continuada. Por outro lado, o indivíduo não tem acesso aos dados por si gerados, nem capacidade de analisá-los em igualdade de circunstâncias. A interação com sistemas de IA afeta a vida do indivíduo. Extrapolando essa premissa, há de se avaliar impactos numa perspectiva societal, tendo em conta instituições, democracias e a sociedade em geral (COM, 2019).

A acumulação de dados recolhidos em espaço público sobre cada indivíduo representa um enorme potencial para o conhecimento da sociedade e a gestão do espaço público para os mais diversos fins, nomeadamente segurança e manutenção da ordem pública. No entanto, os impactos sociais devem ser considerados mediante as transformações que a vigilância massiva introduz na produção social do espaço. Expectativas sobre anonimato e sobre como o espaço público é gerido por meio do exercício de controle, quotidianamente negociado entre vários grupos numa perspectiva dinâmica, podem ficar condicionadas ou mesmo comprometidas com a implementação desse tipo de sistema de vigilância.

Em consequência, há de se considerar que as formas de interagir e as apropriações do espaço podem alterar-se radicalmente e com elas os significados destes espaços para a projeção de identidades coletivas e a promoção de valores democráticos. A perpetuação no tempo pode conduzir ao escalar da conflitualidade entre os vários atores e grupos, mediante a redefinição de relações de poder, e, num outro plano, em que as vulnerabilidades dos sistemas democráticos ficam expostas, abre-se margem para a sedimentação de regimes autoritários.

PROPORCIONALIDADE E EQUIDADE

O princípio da proporcionalidade refere-se à ideia de providenciar um resultado justo, significando que são tidos em consideração o custo, a complexidade e os recursos disponíveis. Proporcionalidade é um princípio geral legal na UE (União Europeia), restringindo as autoridades no exercício dos seus poderes, a fim de ser encontrado um equilíbrio entre os meios usados e os fins a atingir e requerendo o compromisso de que as vantagens na limitação de direitos não são superadas por desvantagens. Isto aplica-se, por exemplo, no âmbito da proteção de dados, considerando-se que apenas

informações adequadas e relevantes para o fim predefinido podem ser recolhidas e processadas, mediante o consentimento do indivíduo.

Na implementação de sistemas de vigilância em espaços públicos, questões de proporcionalidade aplicam-se, no sentido de avaliar e justificar a vigilância massiva, a identificação e a monitorização de comportamentos de potencialmente todos os indivíduos que percorrem o espaço, quando apenas uma ínfima percentagem desses indivíduos está sob observação por motivos judicialmente justificáveis. Especificamente quando estes implicam o uso de tecnologias de IA, como reconhecimento facial, o seu funcionamento fica sujeito à captura e à comparação de imagens, usando uma lista de suspeitos como referência. Isto significa que todos os indivíduos em espaço público se encontram sob escrutínio. Kamgar-Parsi *et al.* (2011) referem como o recurso a tecnologias de reconhecimento facial implica uma abordagem distinta na identificação do indivíduo, pois inclui analisar todos os indivíduos num processo contínuo de validação e exclusão, enquanto o processo de identificação levado a cabo por um humano ocorre por meio do reconhecimento de forma descontínua.

Consequentemente, deverão ser tidos em consideração os impactos para todos os indivíduos, numa perspectiva societal, tendo em conta que estarão igualmente expostos à tecnologia. Um destes impactos relaciona-se com a privacidade.

O anonimato é considerado um valor urbano positivo e mesmo intrínseco ao conceito de vida pública urbana (Simmel, 1973). A expectativa de permanecer anônimo na multidão é uma das condições que define a vida pública e um dos mecanismos para a gestão do exercício de poder em espaço público. É também com base nessa premissa que determinados grupos se sentem acolhidos e protegidos em espaço público, seja porque vivem em situações marginais (não significando necessariamente criminalidade), seja porque é nessa condição que interagem em determinados contextos, como manifestações e assembleias relativas aos mais variados temas dentro das dimensões sociocultural e política.

O princípio da equidade implica a garantia de uma distribuição equitativa e justa dos benefícios e dos custos, bem como de inexistência de enviesamentos injustos, discriminação e estigmatização contra pessoas e grupos (COM, 2019). No entanto, o risco, ainda que inadvertido de exclusão de determinados indivíduos ou grupos por meio da introdução de sistemas de vigilância em espaço público, tem vindo a ser reiterado. A implementa-

ção de sistemas de vigilância não impacta equitativamente sobre todos os indivíduos, na medida em que diferentes indivíduos e diferentes grupos têm formas próprias de se relacionar com o espaço público – veja-se o caso dos sem-abrigo, mencionado anteriormente.

Lyon (2001) refere que a vigilância do espaço tende a exacerbar estereótipos, contribuindo para a perpetuação de sistemas discriminatórios imbuídos culturalmente. O risco prende-se com a possibilidade de as tecnologias com base em IA reforçarem a aptidão para identificar e triar quem deve ser vigiado e quem pode ficar livre desse controle, ao invés de favorecerem a retificação de formas de exclusão social.

Quando falamos de espaços públicos e publicamente acessíveis, a característica central é serem tendencialmente abertos a todos, não obstantes as situações de conflitualidade na projeção de significados e inerentes à produção social do espaço. Espaços vigiados por meio de sistemas que permitem identificar indivíduos em tempo real e de forma sistemática são espaços de elevado escrutínio, comparável ao nível que encontramos em espaço doméstico, nos quais a expectativa é ser reconhecido na sua identidade para poder ter acesso. Assim, existe o risco de restrição de acesso, gerando espaços de exclusão, onde os grupos "indesejáveis" se encontram sob maior escrutínio. Aqui, torna-se incontornável pensar em questões de discriminação e privação de direitos adquiridos, bem como na erosão de valores democráticos, como a diversidade e a tolerância, pervertendo os significados de espaço público como espaço aberto e inclusivo.

Em sistemas de IA, os requisitos de diversidade, não discriminação e equidade estão contemplados para garantir que a tecnologia é confiável (COM, 2019), ou seja, a inclusão e a diversidade têm de estar presentes em todo o ciclo de vida do sistema. Esse requisito relaciona-se com o princípio da equidade e prevê o envolvimento de todas as partes interessadas ao longo do processo e à igualdade de acesso (considerando processos de concessão inclusivos e a igualdade de tratamento).

No entanto, os riscos que os sistemas de IA representam em matéria de privacidade e para o respeito de direitos universais colocam, ainda, vários desafios. Por um lado, como mencionamos, o seu funcionamento depende de dados produzidos por humanos e que, em outra medida, representam a sociedade como um todo, onde se encontram imbuídas informações pessoais, mas também estereótipos culturais. Por sua vez, os processos de automatização no processamento devolvem resultados potencialmente

difíceis de avaliar, tendo em conta a sua complexidade e podem significar a reprodução de preconceitos nas suas decisões.

Ainda sobre privacidade e *big data*, os avanços na regulação da proteção de dados deixam patente a ideia de autodeterminação sobre o tratamento dados pessoais. Westin (1968) defende que os indivíduos, grupos e instituições têm o direito de controlar, editar, gerir e apagar informações sobre si próprios e decidir quando, como e em que nível essas informações podem ser partilhadas com outros. No Regulamento Geral da Proteção de Dados (EU, 2016), é reiterada a relação potencialmente conflitual entre a utilização de dados pessoais e o respeito pelos direitos fundamentais e liberdade individual. É ainda considerada a repercussão da legislação sobre a monitorização de espaços publicamente acessíveis e a proibição do processamento de dados biométricos com a finalidade de identificar indivíduos, contemplando, apesar de tudo, exceções, nomeadamente quando há consentimento explícito por parte do sujeito.

CONSIDERAÇÕES FINAIS

Com o desenvolvimento tecnológico de novas ferramentas de comunicação e informação interpessoal e societal, a copresença deixa de ser condição fundamental na interação. As interações podem ser mediadas por tecnologias que, mediante o registro sistemático e digitalização da informação, facilitam a comunicação e a transmissão de informação, recolhendo, mantendo e processando dados, que podem ser veiculados em cadeias de várias dimensões, numa perspectiva global. Em consequência, o que é considerado informação pessoal e informação pública tem vindo a complexificar-se (Lyon, 2001). A rapidez com que o processo vem ocorrendo nem sempre contemporiza a necessidade de informar e consultar o indivíduo sobre o assunto e de regular a utilização e a reutilização de dados pessoais, nas variadas esferas de aplicação, avaliando os diferentes riscos e refletindo sobre os novos desafios que as tecnologias introduzem na vida quotidiana e para a sociedade em geral.

O uso de tecnologias de reconhecimento facial e identificação biométrica enquadra-se neste panorama. As esferas de implementação são múltiplas. Funcionando de forma semelhante a uma *password* para garantir acessos, no controle de fronteiras, para a validação de presenças e na vigilância do espaço público, a lista de potenciais aplicações para estas tecnologias, que integram componentes baseados em IA, continua a crescer.

Os avanços permitidos pela IA, como aumento da capacidade e rapidez no processamento de dados e aumento da autonomia nos processos e tomadas de decisão, representam novas oportunidades para a sociedade, providenciando ao ser humano ferramentas que podem nomeadamente libertá-lo de determinadas tarefas e apoiar na concessão de soluções para problemas que afetam negativamente a vida e a dignidade humanas, fomentando o melhoramento das condições para o bem comum e o desenvolvimento sustentável (Taddeo; Floridi, 2018; Floridi; Cowls; King; Taddeo, 2020).

No entanto, os impactos potenciais deste tipo de tecnologia não são exclusivamente positivos. Existem riscos associados à implementação, nomeadamente para a preservação de valores estabelecidos considerados positivos e aportam novos desafios no respeito por direitos adquiridos. A complexidade da tecnologia e as mudanças de paradigma em termos de quem está em condições de interpretar informação e quem está capacitado a tomar decisões têm vindo a reforçar a necessidade de supervisão e de intervenção humana, reforçando a posição do humano como detentor do controle sobre a tecnologia.

A reflexão proposta incide sobre o impacto do uso de sistemas de identificação biométrica em tempo real e a distância, em particular tecnologias de reconhecimento facial em espaços públicos. A abordagem parte, no entanto, da premissa de que nem todos os riscos são oriundos da tecnologia em si, mas podem, mediante o seu enorme contributo potencial, ser amplificados. Os impactos da implementação de sistemas de vigilância em espaço público, integrando nomeadamente a identificação de indivíduos a distância e em tempo real, impactam sobre as relações de poder que gerem o funcionamento da vida pública e significados do espaço público, podendo aprofundar assimetrias, onde sai reforçado o papel do Estado e das corporações e expondo grupos vulneráveis a novos constrangimentos no *direito à cidade*.

Diversos atores projetam diferentes narrativas e têm diferentes expectativas sobre quais os limites nas apropriações do espaço e quais os níveis de controle e escrutínio aceitáveis. Estes dependem de valores culturais atribuídos, retratando, por exemplo, a importância destes espaços da cidade no desenvolvimento de atividades relacionadas com a vida pública e na construção de identidades coletivas. Podem, assim, ser entendidos como centrais na exaltação de valores democráticos como a tolerância e a inclusão social.

Neste sentido, o reajuste das relações de poder sobre o espaço, proposto pelo Estado, respaldado na necessidade de manter a segurança e ordem pública, está sujeito à crítica de outros atores. A aceitação de sistemas de vigilância, progressivamente mais intrusivos e omnipresentes, depende, ainda, do nível de confiança que as comunidades depositam nos seus governantes para gerir custos individuais em prole do bem comum. Dependem, também, do quanto estas são sensíveis aos argumentos políticos e de como estes são reflexo das preocupações e problemas contextuais. Isto significa que, em regimes mais opressivos ou em comunidades mais afetadas por problemas de criminalidade, as condições para aceitar a vigilância do espaço público, nestes termos, poderão estar potenciadas, no sentido de que há maior abertura para restringir liberdades individuais e comprometer a privacidade do indivíduo, perante o benefício hipotético de contribuir para o restabelecimento da segurança e da ordem pública.

No espectro oposto, vêm surgindo várias respostas locais e pontuais ou mais globais e sistemáticas, em formas de contestação, propostas de regulação e diretrizes para gerir riscos e aproveitar os potenciais da IA. Como exemplo de uma destas respostas, a Comissão Europeia (COM, 2021) propõe a proibição do uso de tecnologias de identificação biométrica para a identificação de indivíduos a distância e em tempo real (com algumas exceções). Num quadro local, as iniciativas e propostas para a regulação e proibição multiplicam-se, com propostas para banir o uso em várias cidades americanas (São Francisco, Baltimore, Portland, entre outros casos) e Hangzhou ou na perspectiva de regular o uso, como em Buenos Aires e Ningbo[5].

REFERÊNCIAS

AGIER, M. *Antropologia da cidade*: lugares, situações, movimentos. São Paulo: Terceiro Nome, 2011.

ARENDT, H. *The human condition*. Chicago: University of Chicago Press, 1958.

BAUDELAIRE, C. *O spleen de Paris*: pequenos poemas em prosa. Lisboa: Relógio d'Água, v. [1869], 1991.

BENJAMIN, W. *Obras escolhidas III*: Charles Baudelaire, um lírico no auge do capitalismo. São Paulo: Brasiliense, 1991.

5 Informação obtida a partir do repositório criado no âmbito do projeto "AI Localism. The responsible use and design of artificial intelligence at the local level". GovLab. Disponível em: https://ailocalism.org/.

COM. *Ethics guidelines for trustworthy AI*. European Commission: 2019.

COM. *Laying down harmonised rules on artificial intelligence (Artificial Intelligenc Act) and mending certain union legislative acts*. European Commission, 2021.

DAMATTA, R. *A casa & a rua*. Espaço, cidadania, mulher e morte no Brasil. 5. ed. Rio de Janeiro: Rocco Digital, 1997.

DITTON, J.; SHORT, E. Yes, it works, no, it doesn't: comparing the effects of open-street CCTV in two adjacent Scottish Town Centres. In: CLIVE NORRIS, D. W. *Surveillance, crime and social control*. [s.l.]: Routledge, p. 201-223, 2006.

EDPB EDPS. *Joint opinion on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. 2021.

EU. *General Data Protection Regulation* – Regulation (EU) 2016/679. 2016.

FLORIDI, L. et al. AI4People – An ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, v. 28, p. 689-707, 2018.

_____ et al. How to design AI for social good: seven essential factors. *Science and Engineering Ethics*, v. 26, p. 1771-1796, 2020.

FLORINI, A. M. The battle over transparency. In: FLORINI, A. M. *The right to know*: transparency for an open world. New York: Columbia University Press, p. 1-16, 2007.

FONTES, A. C. *As pequenas escalas da rua*. Tese de Doutoramento em Estudos Urbanos. ISCTE-IUL e FCSH-UNL, 2020.

FOUCAULT, M. *Discipline and punish*: the birth of the prison. New York: Random House, 1979.

FYFE, N. R.; BANNISTER, J. "The eyes upon the street": closed-circuit television surveillance and the city. In: FYFE, N. R. *Images of the street*: representation, experience and control in public space. London: Routledge, p. 254-267, 1998.

GEHL, J.; GEMZøE, L. *New city spaces*. The Danish Architectural Press, 2001.

HABERMAS, J. *Espace public*. Paris: Payot, 1978.

HABRAKEN, N. J. *The structure of the ordinary*. Form and Control in the Built Environment. Massachusetts: MIT Press, 1998.

HANNERZ, U. *Exploring the city*. Inquires toward an urban anthropology. New York Chichester: Columbia University Press, 1980.

JACOBS, J. *Death and life of great American cities*. New York: Random House, 1961. [1961].

JINU, C.; SHEEJA, A. A. Face recognition in CCTV systems. *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2019.

KAMGAR-PARSI, B.; LAWSON, W.; KAMGAR-PARSI, B. Toward development of a face recognition system for watchlist surveillance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 33, n. 10, 2011.

KOSKELA, H. "The gaze without eyes": video-surveillance and the changing nature of urban space. *Progress in Human Geography*, v. 24, n. 2, p. 243-265, 2000.

_____. "Cam Era" – The contemporary urban panopticon. *Surveillance & Society*, v. 1, n. 3, p. 292-313, 2003.

LEFEBVRE, H. *The production of space*. Oxford UK e Cambridge USA: Blackwell, 1991.

LYON, D. *Surveillance society*. Monitoring everyday life. Buckingham and Philadelphia: Open University Press, 2001.

MADANIPOUR, A. *Public and private spaces of the city*. Routledge, 2003.

_____. Introduction. In: MADANIPOUR, A. *Whose public space?* International case studies in urban design and development. Abingdon and New York: Routledge, p. 1-15, 2010.

MITCHELL, D. The end of public space: people's park, definitions of the public and democracy. *Annals of the Association of American Geographers*, v. 85, p. 108-33, 1995.

NORRIS, C.; ARMSTRONG, G. CCTV and the social structuring of surveillance. In: NORRIS, C.; WILSON, D. *Surveillance, crime and social control*. Routledge, p. 157-178, 2006.

ROO, M. S. Human activity prediction: early recognition of ongoing activities from streaming videos. *IEEE International Conference on Computer Vision*, Barcelona, 2011.

SENNETT, R. *The fall of public man*: on the social psychology of capitalism. New York: Random House, 1977.

SI, W. et al. Remote identity verification using gait analysis and face recognition. *Wireless Communications and Mobile Computing*, 2020.

SIMMEL, G. A metrópole e a vida mental. In: VELHO, O. *O fenômeno urbano*. Rio de Janeiro: Zahar, 1973. [1950].

SQUIRES, J. Private lives, seculed places: privacy as a polical possibility. *Environment and Planning D: Society and Space*, v. 12, n. 390, 1994.

TADDEO, M.; FLORIDI, L. How AI can be a force for good. *Science*, v. 361, n. 6404, p. 751-752, 2018.

TAYLOR, N. State surveillance and the right to privacy. *Surveillance & Society*, v. 1, n. 1, p. 66-85, 2002.

UNITED NATIONS. *Universal Declaration of Human Rights*. General Assembly Resolution 217 A, 1948.

WESTIN, A. F. Privacy and freedom. *Wash. & Lee L. Rev.*, v. 25, n. 1, 1968.

Sobre a autora e o autor:

Catarina Fontes

Research Associate at Institute for Ethics in Artificial Intelligence – TUM.

Christoph Lütge | *E-mail:* luetge(at)tum.de

Prof. Lütge (*1969) forscht auf dem Gebiet der Wirtschafts- und Unternehmensethik. Er vertritt den Ansatz einer Ordnungsethik, der ethisches Handeln unter den ökonomischen und sozialen Rahmenbedingungen der Globalisierung erforscht. Die Rolle des Wettbewerbs und der von Ordnungen ausgehenden Anreize stehen dabei ebenso im Vordergrund wie die Prüfung ethischer Kategorien auf Angemessenheit. Nach dem Studium der Wirtschaftsinformatik und Philosophie (Promotion 1999) war Prof. Lütge wissenschaftlicher Assistent am Lehrstuhl für Philosophie und Ökonomik der LMU München, wo er sich auch habilitierte (2005). Forschungsaufenthalte führten ihn nach Pittsburgh, San Diego und Venedig. Von 2007 bis 2010 vertrat er Lehrstühle in Witten/Herdecke und Braunschweig und hat seit 2010 den Peter Löscher-Stiftungslehrstuhl für Wirtschaftsethik an der TUM inne.

Artigo convidado.

Internet and the Political Campaigns. The Perspective of the Italian Legislation

Internet e Campagne Elettorali. La Prospettiva Dell'Ordinamento Italiano

ORNELLA SPATARO[1]
Università degli studi di Palermo, Sicilia, Italia.

ABSTRACT: This paper intends to discuss the issue of regulation of social networks as a media space employed by political actors (individuals or organizations) for the formation of public opinion and the construction of consensus in the context of the electoral campaigns: in more synthetic terms, social networks as an instrument of electoral propaganda. The main subject of my essay will concern, therefore, the problems *de iure condito* and *de iure condendo* about the subjection of communication on social media to rules aimed at ensuring a fair comparison between parties, lists, candidates in the electoral competition and, with it, the value of freedom and authenticity of the vote that emanates from article 48 of the Italian Constitution. This substantiates a constitutional directive in favor of a regulatory intervention of the legislator, called upon to achieve an adequate balance between the many rights and interests involved.

KEYWORDS: social media; political communication; electoral campaigns; freedom of speech; right to information; digital populism; liberal democratic theory; free and conscious vote; free and fair competition between political parties.

ABSTRACT: Questo scritto affronta il tema della regolazione dei social network quale spazio mediatico utilizzato dai soggetti politici (individui od organizzazioni) per influenzare la formazione dell'opinione pubblica e costruire il consenso a proprio favore, soprattutto nell'ambito delle campagne elettorali. La questione concerne l'uso dei social come strumento di propaganda elettorale che, per le sue caratteristiche tecniche, sfugge all'applicazione della disciplina dettata per i media tradizionali. L'articolo si soffermerà, dunque, sui problemi *de iure condito* e *de iure condendo* circa l'assoggettabilità della comunicazione via social a regole volte a garantire un corretto confronto tra partiti, liste, candidati nella competizione elettorale e, con esso, quel valore della libertà e genuinità del voto che promana dall'art. 48 Cost. Tali princìpi pongono l'esigenza costituzionale di un intervento regolatorio del legislatore, chiamato a realizzare un adeguato bilanciamento tra i numerosi diritti e

1 Orcid: https://orcid.org/0000-0002-9172-3604

interessi in gioco, e a garantire il più possibile la correttezza della competizione politica come sistema procedurale su cui si fondano le democrazie liberali.

PAROLE CHIAVE: social media; comunicazione politica; campagne elettorali; libertà di espressione; diritto all'informazione; populismo digitale; teoria democratica liberale; voto libero e consapevole; competizione corretta e leale tra i partiti politici.

1 SOCIAL MEDIA E CAMPAGNE ELETTORALI. UN BINOMIO PROBLEMATICO

Il ruolo che le piattaforme social hanno assunto nella sfera pubblica rischia di incidere su alcuni dei valori fondamentali dell'ordinamento democratico. In particolare, tali rischi risultano evidentissimi se si ha riguardo allo svolgimento delle campagne elettorali, che sono prodromiche alla corretta realizzazione di un momento costitutivo di ogni sistema democratico, quello delle consultazioni elettorali. Guardando all'attuale assetto normativo, si cercherà di evidenziare come l'approccio regolativo adottato dall'ordinamento italiano appaia insufficiente a garantire le condizioni procedurali che il costituzionalismo liberale pone a fondamento della democrazia.

La cronaca degli ultimi anni mostra chiaramente quale possa essere l'impatto delle piattaforme *social* sul corretto svolgimento del processo elettorale. Basti pensare alle elezioni presidenziali statunitensi del 2017, in seguito alle quali è emerso il cosiddetto "*Russia gate*", che ha reso evidente dinanzi all'opinione pubblica come l'uso delle piattaforme possa avere effetti distorsivi sulla formazione della volontà elettorale dei cittadini.

In particolare, da alcuni documenti dell'intelligence statunitense[2], oltre che da indagini interne svolte da Facebook (STAMOS, 2017), è emerso il fondato sospetto che la volontà politica degli elettori fosse stata manipolata da specifiche e mirate attività di disinformazione condotte sui *social*, con effetti determinanti sul risultato elettorale: è stato calcolato, infatti, come la diffusione delle *fake news* (SCIORTINO, 2020) sui social

2 Office of the Director Of National Intelligence, *Background to "Assessing Activities and Intentions in Recent US Elections": The Analytic Process and Cyber Incident Attribution*, 6 gennaio 2017.

si fosse concentrata soprattutto negli "*Swing States*", ovvero negli "Stati in bilico" che sono decisivi per la competizione elettorale[3] (BESSI, FERRARA, 2017; HOWARD, KOLLANYI, BRADSHAW, NEUDERT, 2017).

L'esempio del *Russia gate* pone due ordini di problemi. Innanzitutto si pone la questione della possibilità che Stati esteri usino le piattaforme per influenzare le elezioni che si svolgono in altri Paesi, condizionandone l'esito per scopi di politica internazionale. In termini più ampi il tema concerne il ruolo delle piattaforme rispetto al processo elettorale, e la diffusione di notizie offensive e tendenziose sui *social*, che possono condizionare la formazione e l'espressione della volontà politica dei soggetti.

Ciò si inserisce nel contesto delle trasformazioni che la rivoluzione informatica ha impresso alla società, con ricadute importanti che riguardano l'ambito della politica. L'incremento progressivo della capacità computazionale, con il passaggio dal cosiddetto web 1.0 al web 2.09, abbinato ad un andamento discendente della curva dei costi, ha favorito l'espansione di un sistema che assicura una inedita posizione di potere ad un oligopolio di piattaforme. Esse si trovano, infatti, a gestire un enorme flusso di informazioni, relativo a circa un quarto della popolazione mondiale che ha accesso ad internet (FLORIDI 2017).

E' noto che gli utenti delle piattaforme, nel momento in cui vi accedono, lasciano tracce, che vengono raccolte e combinate attraverso processi di *data mining* (COLONNA, 2013; AMPOFO, COLLISTER, O'LOGHLIN, CHADWICK, 2015); ciò attiva un'enorme accumulazione di conoscenza, che, per le piattaforme che la detengono, si traduce in un potere che è tanto più forte e pervasivo quanto più i processi che ne favoriscono la formazione si svolgono all'insaputa degli utenti (ULBRICHT, VON GRAFESTEIN, 2016). La conoscenza degli utenti della rete genera una sorveglianza costante (ZUBOFF, 2015), su cui si fonda la possibilità per le piattaforme di esercitare il controllo sulle preferenze dei soggetti attraverso la progettazione degli algoritmi (KROLL, HUEY, BAROCAS, 2017; ANNANY, 2016). I *social media* sono, infatti, società che fondano il loro successo economico sulla vendita dei dati degli utenti nel mercato delle inserzioni pubblicitarie; si tratta dunque di processi che originariamente sono stati pensati e applicati alle strategie commerciali, e che sono stati successivamente trasferiti all'ambito

3 V., In particolare, le ricerche condotte nell'ambito del *Computational Propaganda Research Project* dell'Oxford Internet Institute Computational Propaganda Research Project, *http://comprop.oii.ox.ac.uk*.

della comunicazione politica: le forze politiche, come le imprese per la vendita dei propri prodotti, sfruttano la conoscenza della popolazione, acquisita tramite la raccolta e il trattamento dei dati, per decidere a chi, in che modo e cosa comunicare al fine di massimizzare le probabilità di un esito elettorale favorevole.

I *social media* sono perfettamente strumentali alla realizzazione di questi obiettivi, perché sono in grado di raccogliere informazioni estremamente dettagliate sulle caratteristiche dei propri utenti, consentendo la formulazione di messaggi esattamente calibrati in relazione alle preferenze dei destinatari. Sulla base di questo, le piattaforme sono in grado di comunicare con un *target* ridotto e selezionato di soggetti: Facebook, ad esempio, permette di personalizzare le proprie inserzioni, selezionando il pubblico che le riceve, e realizzando messaggi che saranno visualizzati solo dagli utenti individuati in base agli specifici criteri selezionati dall'inserzionista (DELANY, 2014). Utilizzando queste caratteristiche tecniche, ad esempio, lo staff elettorale di Trump poteva poteva diffondere ogni giorno circa 50.000 diverse versioni dei messaggi politici, addirittura rivolgendosi a poche decine di utenti in una particolare circoscrizione (BECKETT, 2017).

E' evidente che si tratta di meccanismi che hanno potenzialità estremamente manipolative: la raccolta e lo sfruttamento economico delle informazioni tratte dalla rete ha consentito alle compagnie che gestiscono le piattaforme di generare enormi profitti, sfruttando le dinamiche di mercato. Se però i medesimi meccanismi vengono utilizzati per generare consenso politico, è chiaro che si pongono problemi rilevantissimi sul piano della correttezza della deliberazione pubblica e della compatibilità con i princìpi democratici. Nel momento stesso in cui l'utente accede alle piattaforme per acquisire informazioni, effettuando ricerche, diventa egli stesso oggetto di informazione, cedendo i dati che riguardano i propri interessi, i propri gusti, le proprie esigenze, le proprie abitudini, le relazioni sociali. Questo consente alle piattaforme di ricostruire la consistenza e l'orientamento dell'elettorato, combinando dati che possono essere offerti ai soggetti politici per precostituirsi un vantaggio elettorale, e per orientare i temi della comunicazione in modo da intercettare le preferenze dell'elettorato.

Non ci si può non chiedere se e quanto questo scenario sia compatibile con la teoria democratica liberale, che presuppone la sussistenza di un'opinione pubblica autonoma, capace di esprimere attraverso le elezioni le proprie preferenze, e che pone come requisito di sistema lo svolgimento di una competizione libera e leale tra i partiti politici.

Questi principi sono riconosciuti non solo dagli ordinamenti statali, ma anche a livello sovranazionale[4]: secondo la Corte europea dei diritti dell'uomo[5], la democrazia «si nutre della libertà di espressione»[6], che costituisce uno dei fondamenti essenziali di una società democratica e una delle condizioni principali del suo progresso[7]. Per altro verso, la medesima Corte ha affermato l'esigenza di tutelare il correlativo diritto del pubblico di ricevere informazioni[8].

Guardando all'ordinamento italiano, anche la Corte costituzionale ha riconosciuto l'importanza del diritto all'informazione[9]: il giudice delle leggi ha da tempo affermato che quello all'informazione è «uno tra i principi caratterizzanti del vigente ordinamento democratico», poiché svolge un ruolo fondamentale per la «formazione di una pubblica opinione avvertita e consapevole»[10], e che il regime democratico «implica pluralità di fonti di informazione, libero accesso alle medesime, assenza di ingiustificati ostacoli legali, anche temporanei, alla circolazione delle notizie e delle idee»[11].

La stessa importanza è stata riconosciuta alla funzione svolta dalla comunicazione politica[12]: tanto il quadro giuridico convenzionale, quanto

4 Nell'ambito del Consiglio d'Europa, tali principi sono stati elaborati da fonti si *soft law*. Si vedano, ad esempio, la raccomandazione n. R(99)1 del Comitato dei Ministri agli Stati membri volta a promuovere il pluralismo dei media; la raccomandazione Rec(2003)9 del Comitato dei Ministri agli Stati membri sulle misure volte a promuovere il contributo democratico e sociale della radiodiffusione digitale; la raccomandazione CM/Rec(2007)2 del Comitato dei Ministri agli Stati membri sul pluralismo dei media e la diversità del contenuto dei media; deve inoltre essere segnalato il *Code of Good Practice in Electoral Matters* elaborato dalla *European Commission for Democracy through Law*.

5 I principi sono richiamati dalla Grande Camera della Corte di Strasburgo nella sentenza *Centro Europa 7 s.r.l.* e *Di Stefano c. Italia*, n. 38433/09, 7 giugno 2012, parr. 129-134.

6 Corte europea dei diritti dell'uomo, *Manole* e *altri c. Moldova,* n. 13936/02, par. 95; Partito socialista e altri c. Turchia, 25 maggio 1998, parr. 41, 45, 47.

7 Corte europea dei diritti dell'uomo, *Lingens c. Austria*, 8 luglio 1986, par. 41. 22 Corte europea dei diritti dell'uomo, *Handyside c. Regno Unito*, 7 dicembre 1976, par. 49; Lingens c. Austria, parr. 41-42.

8 22 Corte europea dei diritti dell'uomo, Handyside c. Regno Unito, 7 dicembre 1976, par. 49; Lingens c. Austria, parr. 41-42.

9 Corte cost., sentenza n. 826/1988, *considerato in diritto* nn. 9-11; sentenza n. 112 /1993, *considerato in diritto* n. 7; sentenza n. 420/1994, *considerato in diritto* nn. 14.2, 14.3.

10 Corte cost., sentenza 15 giugno 1972, n. 105, *considerato in diritto*, n. 3.

11 Ivi, considerato in diritto, n. 4.

12 Nel caso della comunicazione politica i contenuti sono prodotti direttamente da parte dei soggetti politici (messaggi autogestiti, tribune politiche); l'informazione, invece, ha ad oggetto contenuti creati dai giornalisti, che, ovviamente, possono anche riguardare i politici. In Italia, l'art. 5 l. 22 febbraio 2000, n. 28 prevede che, in campagna elettorale, anche i programmi d'informazione siano soggetti al rispetto di specifici criteri definiti dalla Commissione parlamentare per l'indirizzo generale e la vigilanza dei servizi radiotelevisivi e dall'Autorità per le garanzie nelle comunicazioni «al fine di garantire la parità di trattamento, l'obiettività, la completezza e l'imparzialità dell'informazione». La legge sancisce il divieto di «fornire, anche in forma indiretta, indicazioni di voto o manifestare le proprie preferenze di voto» (co. 2), assoggettando conduttori e registi dei programmi di informazione all'obbligo di tenere «un comportamento corretto ed imparziale nella gestione del programma, così da non esercitare, anche in forma surrettizia, influenza sulle libere scelte degli elettori» (co. 3).

l'ordinamento nazionale impongono allo Stato l'esigenza di adottare particolari garanzie per la fase delle elezioni. Lo svolgimento delle campagne elettorali rientra nel campo di applicazione del diritto sancito all'art. 3 del Protocollo I della CEDU, per cui sullo Stato grava l'obbligo positivo di predisporre gli strumenti legislativi e amministrativi idonei a garantire un pluralismo effettivo, nonché di regolare le procedure elettorali «*under conditions which will ensure the free expression of the opinion of the people in the choice of the legislature*»[13].

La giurisprudenza della Corte costituzionale italiana individua la *ratio* sottesa alla disciplina in materia di propaganda e comunicazione politica nella necessità di evitare che, in una situazione come quella della campagna elettorale, momento di «concomitante e più intensa partecipazione di partiti e di cittadini alla propaganda politica», vengano a determinarsi degli squilibri che potrebbero derivare da «situazioni economiche di svantaggio o politiche di minoranza»[14].

Nel rispetto di tali princìpi il legislatore ha disciplinato il rapporto tra la comunicazione politica e i media tradizionali allo scopo di tutelare il più possibile la correttezza dell'informazione, garantendo il pluralismo, e collegando al potere editoriale dei soggetti professionali particolari forme di responsabilità. Ciò nella consapevolezza che disciplinando il potere degli intermediari dell'informazione, e garantendo la correttezza della comunicazione politica e della competizione tra i partiti, si può, anche se indirettamente, tutelare la sfera cognitiva degli elettori[15].

E' evidente, però, che l'effetto distorsivo dei social sulla formazione dell'opinione pubblica, in particolare nel momento della campagna elettorale, sfugge a questo tipo di regolamentazione, ponendo problemi nuovi agli ordinamenti giuridici.

13 Corte europea dei diritti dell'uomo, *The Communist Party of Russia and Others v. Russia,* n. 29400/05, 19 giugno 2012, parr. 107-108, 118, 134; *Yumak and Sadak v. Turkey,* n. 10226/03, 8 luglio 2008.

14 Corte cost., sentenza 16 giugno 1064, n. 48.

15 La disciplina italiana è dettata dalla l. 4 aprile 1956, n. 212, *Norme per la disciplina della propaganda elettorale*, dalla l. 10 dicembre 1993, n. 515, *Disciplina delle campagne elettorali per l'elezione alla Camera dei deputati e al Senato della Repubblica*, dalla l. 22 febbraio 2000, n. 28, *Disposizioni per la parità di accesso ai mezzi di informazione durante le campagne elettorali e referendarie e per la comunicazione politica*, come modificata dalla l. 6 novembre 2003, n. 313, *Disposizioni per l'attuazione del principio del pluralismo nella programmazione delle emittenti radiofoniche e televisive locali*.

2 LA COMUNICAZIONE POLITICA SUI SOCIAL TRA LIBERTÀ DI ESPRESSIONE E DIRITTO ALL'INFORMAZIONE. IL POPULISMO DIGITALE

Per affrontare tali questioni il diritto dovrebbe confrontarsi di più con la *Computational Social Science* (LAZER, PENTLAND, ADAMIC, 2009; FARO, LETTIERI, 2013). Alcune ricerche condotte in quest'ambito hanno analizzato, in particolare, gli effetti che gli intermediari digitali possono esercitare sulle consultazioni elettorali (REED, 2013). Secondo alcune elaborazioni l'ordine di apparizione dei risultati di ricerca su *Google* può influenzare le preferenze di voto degli elettori indecisi fino al 20% (EPSTAIN, ROBERTSON, 2015). Un celebre studio del 2014 (BOND, FARISS, JONES, 2014), invece, ha evidenziato la corrispondenza tra la diffusione del messaggio "*I vote*" su *Facebook* durante le elezioni statunitensi del 2010 e l'affluenza ai seggi elettorali. Secondo lo studio, il messaggio avrebbe determinato l'affluenza al voto di circa 60.000 elettori, e indirettamente, attraverso il cosiddetto *social contagion*, avrebbe prodotto un incremento di altri 280.000 voti. Nonostante tale cifra rappresenti una percentuale molto bassa rispetto alla totalità dell'elettorato degli Stati Uniti (che alle elezioni statunitensi del 2010 era calcolato in 236 milioni di elettori), è stato evidenziato che il voto di contingenti minimi di elettori indecisi o di elettori di un determinato collegio spesso assume un peso decisivo per l'esito delle consultazioni (TALBOT, 2012)[16].

Ma non è solo un problema quantitativo: gli strumenti digitali hanno impresso cambiamenti cruciali nella qualità del rapporto di ciascun soggetto con la politica. A differenza dei media tradizionali, Internet consente a ciascuno di comunicare con tutti gli altri, interagendo tramite siti, portali, o piattaforme *social*; inoltre, ognuno ritiene di fare le proprie scelte sulla base delle idee personali acquisite tramite l'informazione che circola nel web. Ne consegue che il soggetto non ha più bisogno dei mediatori istituzionali (Parlamento, partiti, sindacati) oppure informali (giornali, telegiornali, gruppi d'opinione, Chiese, associazioni della "società civile"), perché ognuno si informa autonomamente e interagisce direttamente con i soggetti della politica attraverso internet. Naturalmente tutto ciò determina esiti politici di grandissima rilevanza, che favoriscono l'emergere del populismo mediatico, e cambiano la struttura della società politica.

16 Si pensi alle elezioni statunitensi del 2000, quando George Bush vinse su Al Gore per 537 voti in più in Florida.

I *social media* non sono semplicemente strumenti, ma, piuttosto, sono elementi costitutivi dell'ambiente stesso della democrazia. Secondo la prima legge della tecnologia di Melvin Kranzberg (KRANZBERG, 1986), «La tecnologia non è né buona né cattiva, ma non è neanche neutrale»; questa "legge" vale anche per le tecnologie mediatiche, poiché nessuna di esse è neutrale: sia singolarmente, sia nel loro complesso, formano diversi ambienti, cui i diversi sistemi politici si sono diversamente adeguati.

La comunicazione tramite il web tende addirittura ad assorbire interamente la dimensione politica (FOSTER, 2012; KOHL, 2012; NAPOLI, 2013; CUNIBERTI, 2015): venendo meno la narrazione di temi e discorsi politici che passava attraverso la mediazione dei partiti, delle istituzioni o dei giornalisti, le piattaforme si sostituiscono alla politica stessa, proponendosi come strumento di interazione immediata con i soggetti. Ne deriva come conseguenza che la dimensione virtuale è sempre più condizionante rispetto alle decisioni politiche, e, per altro verso, che nel mondo delle piatteforme si innescano i conflitti che tendono al dominio dell'influenza sulle persone. Il rapporto dei partiti politici con i soggetti è dunque monopolizzato dal *web*, ed è nel mondo virtuale che i partiti stessi operano per catturare l'attenzione ed il consenso dei consociati.

Sul piano delle campagne elettorali l'impatto del *web* non riguarda solo la struttura della comunicazione politica, ma incide sulla selezione dei temi e dei protagonisti: il dibattito sulle piattaforme privilegia, infatti, quegli argomenti che sono più idonei ad esercitare una facile ed immediata presa sull'attenzione dei soggetti, secondo gli schemi tipici del populismo (ad esempio i temi dell'immigrazione e della sicurezza, che generano paura nei consociati e consenso verso chi si propone come il difensore degli interessi nazionali, o come giustiziere). Inoltre, si affermano come protagonisti del dibattito i leaders che strumentalizzano gli umori della platea in ascolto, captandone le ondate emotive e lanciando messaggi di corrispondenza con queste ultime (SUNSTEIN, 2006; PARISER, 2011; ZUIDERVEEN BORGESIUS, D. TRILLING, J. MÖLLER et al., 2016).

Questo, abbinato agli strumenti offerti dalle piattaforme che, con i *like*, consentono di verificare in tempo reale il gradimento del pubblico, determina una vera e propria deriva comunicativa, in cui i partiti politici misurano continuamente il proprio successo, con la conseguenza che l'esigenza di comunicazione con il pubblico e di identificazione con i suoi umori prevale sulla politica in senso stretto e sulla problematizzazione delle relative questioni. La battaglia delle idee è destinata a recedere dinanzi

all'esigenza di comunicare messaggi che siano il più possibile graditi al pubblico: Internet si presta a fungere da tecnologia comunicativa ad alto tasso di semplificazione, e, per questo, si pone in facile simbiosi con il populismo, favorendo l'ascesa dei movimenti dell'antipolitica, che mettono in discussione l'assetto liberal-democratico degli Stati.

Il legame tra comunicazione politica nella rete e populismo pone problemi giuridici di sistema, perché se la comunicazione politica cambia così profondamente, non è pensabile che la regolazione normativa resti ferma al passato. Il dibattito giuridico ha delineato chiaramente i rischi che derivano, sia per la tutela dei diritti individuali, sia per il buon funzionamento del sistema democratico, dalla diffusione di un modello politico "sondocratico" (RODOTA', 2004), che si affida in misura sempre maggiore alla sorveglianza e a strategie di comunicazione basate sulla persuasione subliminale (ZOLO, 2001). Tali pratiche infatti, mettendo a rischio la capacità di controllare, filtrare e interpretare razionalmente le informazioni, incidono fortemente sull'autonomia cognitiva del soggetto (ZOLO, 2016).

Ne risulta messa in crisi la stessa teoria liberale che presuppone, quale condizione procedurale della democrazia, una libera e corretta competizione tra i partiti, funzionale alla formazione di un'opinione pubblica autonoma, e, dunque, all'esercizio libero e consapevole del diritto di voto.

La comunicazione politica sui social è orientata a cogliere gli umori degli utenti, per intercettarne il consenso, misurato continuamente tramite i "*like*" in una sorta di continuo test elettorale; ciò fa sì che la propaganda sia permanente: da un lato essa è modellata sulle onde emotive che emergono dalla società; dall'altro, attraverso la persuasione subliminale, essa tende costantemente ad orientare l'opinione pubblica. Nelle fasi di campagna elettorale vera e propria, invece, l'obiettivo della comunicazione politica è quello di acquisire il consenso per spostare il risultato elettorale, anche ricorrendo alla diffusione di informazioni inattendibili, allo *hate speech*, alle *fake news*.

Dunque, cambia radicalmente il modo in cui i partiti organizzano le campagne elettorali (HERSCH, 2015; BENNETT, 2016; KREISS, 2016; GOODMAN, LABO, TAMBINI, MOORE, 2017; CHESTER, MONTGOMERY, 2017): i partiti populisti antisistema utilizzano la comunicazione digitale in modo unidirezionale, proponendo messaggi che vengono sottratti al dibattito giornalistico. Questa forma di comunicazione politica procede costruendo temi e narrazioni che vengono presentate come inattaccabili,

e che riescono a sfruttare, alimentandoli ulteriormente, la sfiducia e il malcontento degli elettori nei confronti dei partiti tradizionali. Inoltre, i cittadini coltivano l'illusione di essere soggetti attivi attraverso l'interazione sui *social*, quando invece, molto spesso, essi sono passivamente esposti all'opera persuasiva degli autori di *blog* o dei siti di informazione che si avvalgono di veri e propri professionisti della comunicazione *social*, e che orientano il dibattito all'interno dei *networks*. Ciò evidenzia come la rete presenti un problema di democraticità, poiché, in essa si ripropongono e si amplificano le disuguaglianze in termini di conoscenza e di *know how*.

Basti pensare al risultato di alcuni appuntamenti elettorali degli ultimi quattro anni, che hanno cambiato profondamente l'assetto politico internazionale: il referendum inglese sulla *Brexit*, l'elezione di Donald Trump, e l'esito delle ultime elezioni politiche in Italia, con la vittoria dei partiti antisistema. Da questi esempi risulta evidente che l'affermazione dei partiti populisti, che si propongono come alternativa al sistema democratico rappresentativo, è stata favorita da campagne elettorali condotte attraverso le piattaforme telematiche: la comunicazione digitale ha contribuito al diffondersi di movimenti di opinione che hanno strumentalizzato la polarizzazione ideologica contro le *elités* della politica tradizionale.

Tutto ciò rende evidente come la questione della regolazione della comunicazione politica sui social nel periodo di campagna elettorale sia cruciale: è in questa fase che l'elettore si forma il proprio convincimento sull'esercizio del diritto di voto, scegliendo tra elenchi definiti di liste e di candidati in vista di una consultazione elettorale determinata.

Come si è visto, l'impatto dei social determina rilevantissime trasformazioni della propaganda politica, cambiandone non solo le forme, ma anche i contenuti, e favorendo l'ascesa dei partiti politici e dei leaders più inclini a sfruttare le nuove modalità della comunicazione digitale. Mentre, per altro verso, l'iperpartecipazione dei soggetti nella rete, determinandone l'esposizione alle tecniche di persuasione tipiche dei *social*, determina un incremento solo apparente della base deliberativa dei processi politici e della cittadinanza attiva. Tutto questo, in un periodo storico di grave indebolimento dei sistemi democratici di governo, e di forte ascesa del potere delle *corporations* digitali, rende impellente il problema di individuare quale sia la regolamentazione possibile delle nuove pratiche di comunicazione politica, soprattutto nelle fasi di campagna elettorale, onde attenuare gli effetti più gravemente distorsivi della rappresentanza politica democratica, garantendo la correttezza del confronto tra partiti, liste e

candidati in modo tale da assicurare la libertà e la genuinità del voto ai sensi dell'art. 48 della Costituzione italiana.

3 SOCIAL MEDIA, INFORMAZIONE E PLURALISMO. I TENTATIVI DI REGOLAZIONE DELLE CAMPAGNE ELETTORALI SUI SOCIAL NELL'ORDINAMENTO ITALIANO

Il legislatore italiano, con la legge n. 28 del 2000, ha disciplinato la comunicazione politica sui media tradizionali[17], con l'obiettivo di favorire il più possibile l'imparzialità e la correttezza dell'informazione, sia nel periodo della campagna elettorale che al di fuori di esso. La suddetta legge, però, non è idonea ad estendere e il suo ambito di applicazione ai *social networks*. Sussiste pertanto nell'ordinamento unna grave lacuna in materia di regolamentazione della comunicazione politica sui media digitali, con il risultato che il potere di influenza delle piattaforme non è controbilanciato da alcun tipo di controllo.

La legge n. 28 del 2000 predispone strumenti normativi atti ad assicurare la parità di *chance* tra i candidati e i partiti nelle competizioni elettorali, attraverso un'equa distribuzione degli spazi della comunicazione radiotelevisiva, prevedendo che tali spazi siano accessibili a titolo gratuito; al contrario, l'assenza di una disciplina specifica fa sì che la comunicazione politica sui *social* avvenga alla stregua di qualsiasi annuncio pubblicitario (DE MINICO, 2019). Non sussiste alcun limite o obbligo legislativo che impedisca alle piattaforme di praticare trattamenti differenziati nei confronti dei diversi soggetti politici, sul piano delle tariffe come su quello dell'accesso ai servizi, con la conseguenza che trattamenti deteriori o migliori possono essere riservati a certi partiti piuttosto che ad altri nell'ambito delle campagne elettorali.

Dinanzi ad una simile situazione si pone l'esigenza di estendere anche ai media digitali i princìpi che sovrintendono alla disciplina delle campagne elettorali (AVVISATI, 2014; MICONI, 2016; TAURO, 2016; CARAVITA, 2019). Ciò però non è un'operazione esente da criticità. Se, infatti, potrebbero applicarsi anche ai *social* media, con i necessari adattamenti, le norme relative alla riconoscibilità dei messaggi di propaganda e quelle sul

17 Nell'ordinamento italiano anche le testate giornalistiche sono soggette a regolamentazione nel periodo compreso tra la convocazione dei comizi elettorali e il penultimo giorno prima delle elezioni. Ai sensi dell'art. 7, l. 28/2000, onde garantire ai soggetti politici la parità di accesso, gli editori che intendano diffondere messaggi politici elettorali attraverso i loro quotidiani o periodici sono tenuti a darne tempestivo avviso sulle testate secondo le modalità fissate dall'AGCOM.

finanziamento della campagna elettorale, lo stesso non può dirsi per altri aspetti della relativa disciplina. L'estensione delle disposizioni della legge sulla *par condicio* alle piattaforme *social* pone, infatti, una serie di problemi non soltanto di adattamento, bensì, ancor prima, di configurabilità di tali soggetti quali esercenti un servizio pubblico[18].

Sul versante delle spese è ipotizzabile l'applicazione della disciplina sulle campagne elettorali, che prevede tetti massimi di spesa, che potrebbero comportare la previsione di tetti pubblicitari per le piattaforme e di obblighi di trasparenza incombenti su finanziatori e finanziamenti. La l. n. 515 del 1993 prevede limiti alle spese elettorali dei candidati, unitamente all'obbligo in capo a ciascun candidato di individuare un proprio mandatario elettorale (obbligato a tenere un'accurata rendicontazione) e a specifici limiti di spesa nei confronti dei partiti e dei movimenti politici. Anche in questo caso, però, non mancano le criticità: le spese per le campagne elettorali *on line* sono difficilmente tracciabili; le schede di rendicontazione che il mandatario elettorale è tenuto a predisporre, del resto, non prevedono la casella delle spese sostenute per la campagna digitale. La possibilità di effettuare pubblicità *on line* senza indicare il mandatario, evidentemente, rende ancor più complicato controllare se effettivamente tutti i candidati si siano attenuti ai limiti di spesa ammissibili in sede di campagna elettorale.

Gli *internet service provider* non possono essere assoggettati a obblighi di controllo generalizzato sulle informazioni immesse in rete dagli utenti, vietati dalla direttiva 2000/31/CE. I *social* nascono infatti non come enti editoriali, ma come piattaforme *free speech*; di qui l'impossibilità di un intervento preventivo per impedire la diffusione in rete di forme di comunicazione politica in ipotesi lesive del principio della *par condicio*.

Ciò, unitamente alle caratteristiche strutturali che differenziano i nuovi media, rende difficilmente applicabili ai *social* le disposizioni che riguardano i profili della parità di accesso e l'equa distribuzione degli spazi della comunicazione politica. Il principale ostacolo, a questo riguardo, deriva dalla personalizzazione delle informazioni che le piattaforme indirizzano ai singoli utenti: mentre gli utenti dei media tradizionali fruiscono di contenuti che indistintamente si rivolgono ad un'*audience* più o meno ampia, nei *social* ciascuno è destinatario di informazioni specificamente calibrate sulla

18 Corte cost., sentenza n. 155/2002, secondo cui la diffusione radio-TV ha "*carattere di preminente interesse generale*".

base dei dati che ha rilasciato, ed è ovvio che risulta estremamente difficile pensare una disciplina capace di tenere conto degli effetti della *selective exposure*, che sono condizionati dalle scelte singolarmente compiute dagli utenti delle piattaforme.

Una possibile regolamentazione dovrebbe semmai essere anticipata al momento in cui viene progettato l'algoritmo, intervenendo sulla cosiddetta "personalizzazione implicita" (ZUIDERVEEN BORGESIUS, TRILLING, MÖLLER, 2016; PARISER, 2011), ovvero imponendo alle piattaforme l'obbligo di costruire l'algoritmo in modo da assicurare che gli utenti entrino in contatto con fonti di informazione plurali e con contenuti di orientamento diverso rispetto a quello che risulti dalla personalizzazione esplicita. Simili proposte però risentono del fatto che il dibattito sull'effettiva incidenza della personalizzazione esplicita e implicita sul flusso di informazioni attraverso i *social* è ancora fermo ad uno stadio preliminare e solo teorico: le ricerche empiriche condotte dagli studiosi di scienze della comunicazione sono poche, e hanno condotto a risultati poco significativi (LUMB, 2015). Non vi possono essere dubbi sulla circostanza che qualsiasi tentativo di regolare il potere delle piattaforme presuppone l'analisi del modo in cui gli algoritmi su cui esse si basano sfruttano il trattamento dei dati degli utenti; tuttavia, come si è detto, questo è un terreno ancora tutto da esplorare.

Tornando al caso italiano, l'Autorità garante delle comunicazioni, in occasione delle campagne elettorali, è competente a deliberare una serie di disposizioni dirette a regolare l'attività di comunicazione politica sulle emittenti radiofoniche e televisive private (competenza che, per il sistema radio-televisivo pubblico, spetta alla Commissione parlamentare di vigilanza). Esercitando tale competenza in occasione delle elezioni europee e amministrative del 26 maggio 2019, la suddetta Autorità, per la prima volta, ha introdotto alcune previsioni orientate alla garanzia del pluralismo e per la condivisione di video sui *social network* (delibere n. 94 e 109 del 2019).

L'AGCom ha altresì raccomandato la trasparenza dei messaggi pubblicitari trasmessi attraverso i *social network*; ha ritenuto che il divieto di pubblicazione dei sondaggi nei giorni precedenti le elezioni, sancito dall'art. 8 della legge n. 28 del 2000, debba essere applicato a tutti i mezzi di informazione, e ha invitato tutti gli attori del mondo digitale ad osservare la regola sul silenzio nel giorno antecedente alla consultazione elettorale, sancita dall'art. 9 della legge n. 212 del 1956.

Circa le delibere le 2019, si tratta, in realtà, di disposizioni che non hanno valore cogente, come si evince dalla lettura dell'unica disposizione (l'art. 26)[19] che compone il nuovo titolo VI di questi due regolamenti; esse non sono nemmeno inquadrabili nell'ambito del *soft law*, ma si pongono come norme che meramente invitano le parti interessate ad una sostanziale autoregolamentazione. Un ruolo importante viene attribuito al Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione[20] sulle piattaforme digitali che opera presso l'Autorità, cui partecipano i principali operatori del settore (Google, Facebook, Twitter), e che, in occasione delle elezioni politiche del 2018 aveva già adottato apposite Linee guida, inquadrabili come norme di *soft law*, per la parità di accesso alle piattaforme *online* durante la campagna elettorale[21]. La prospettiva è quella di estendere i princìpi della legge n. 28 del 2000 sulla disciplina delle campagne elettorali alla comunicazione politica sui social, sempre nell'ottica della autoregolamentazione[22], con l'assunzione di impegni da parte delle piattaforme *social* e dei soggetti politici (definiti ai sensi della legge del 2000), i quali sono destinatari contestualmente di tutele (la parità di accesso) e di raccomandazioni (per fare un esempio che ha fatto molto discutere: il silenzio elettorale). Per il resto, entrano in gioco, con efficacia vincolante, spezzoni relativi ad altre discipline di settore, come la normativa sull' *e-commerce* e sulla *privacy*.

19 Art. 26 (Tutela del pluralismo sulle piattaforme di condivisione di video): «1. Nell'ambito del Tavolo tecnico per la garanzia del pluralismo e della correttezza dell'informazione sulle piattaforme digitali è assunta ogni utile iniziativa al fine di promuovere l'adozione condivisa di misure di contrasto ai fenomeni di disinformazione e lesione del pluralismo informativo online. L'Autorità promuove, mediante procedure di autoregolamentazione, l'adozione da parte dei fornitori di piattaforme di condivisione di video di misure volte a contrastare la diffusione in rete, e in particolare sui social media, di contenuti in violazione dei principi sanciti a tutela del pluralismo dell'informazione e della correttezza e trasparenza delle notizie e dei messaggi veicolati. 3. Le piattaforme si impegnano ad assicurare il rispetto dei divieti sanciti dalla disciplina legislativa e regolamentare in materia di comunicazione e diffusione dei sondaggi».

20 Il Tavolo tecnico è stato istituito con istituito con la delibera n. 423/17/CONS. 10 Le "Linee guida per la parità di accesso alle piattaforme online durante la campagna elettorale per le elezioni politiche 2018" sono state adottate il 1º febbraio 2018.

21 Le linee guida dell'Autorità sembrano voler favorire la tempestività della reazione di fronte a comportamenti che risulterebbero illeciti, come la diffusione di messaggi o videomessaggi con contenuti illeciti, lesivi dell'onore e della reputazione di altri candidati o che, con montature artefatte di interventi o dichiarazioni di un soggetto, attribuiscano a questi false affermazioni o posizioni non rispondenti al vero.

22 La strategia fondata sulla promozione della autoregolamentazione si riscontra anche a livello sovranazionale. A seguito di iniziative assunte dalla Commissione europea, alcuni gestori di piattaforme (come Google, Facebook, Twitter e Mozilla), nonché alcune associazioni che rappresentano il settore pubblicitario, hanno firmato nell'ottobre 2018 un *Codice di buone pratiche contro la disinformazione* e alcune piattaforme (come Facebook e Twitter) hanno integrato di conseguenza i propri codici di condotta. Inoltre, in occasione delle elezioni britanniche del 2019 Facebook ha deciso di limitare ulteriormente profilazione e *microtargeting*, mentre Twitter ha annunciato di rinunciare alla pubblicità politica a pagamento.

A parte questi frammenti normativi, l'approccio dell'Autorità garante è fondato, come si è visto, sull'autoregolamentazione, e ciò deriva dalla difficoltà dell'ordinamento ad elaborare norme cogenti in una materia, come quella della comunicazione politica sui *social*, che coinvolge non solo le piattaforme ed i loro gestori, ma anche gli utenti privati. E' significativa la vicenda della chiusura del profilo di Casa Pound su Facebook, nella quale, prima della decisione del Tribunale di Roma che infine ne ha ordinato il ripristino[23], una soluzione pratica era stata trovata dall'associazione interessata grazie alla "ospitalità" offerta da un *social network* avente sede nella Federazione russa (MELZI D'ERIL, VIGEVANI, 2019; GRASSO, 2020). Il Tribunale di Roma, peraltro, non ha affrontato affatto la questione della violazione dei princìpi e delle regole costituzionali coinvolti, incentrandosi, piuttosto, sull'esigenza di assicurare, prima di tutto, le ragioni del pluralismo politico BIN, 2019; GOLIA, BEHRING, 2020), in nome delle quali ha disposto la riattivazione della pagina Facebook che era stata oscurata. In un passaggio dell'ordinanza si legge che «il rilievo preminente assunto dal servizio di Facebook (o di altri social network ad esso collegati) con riferimento all'attuazione di principi cardine essenziali dell'ordinamento come quello del pluralismo dei partiti politici (49 Cost.), al punto che il soggetto che non è presente su Facebook è di fatto escluso (o fortemente limitato) dal dibattito politico italiano, come testimoniato dal fatto che la quasi totalità degli esponenti politici italiani quotidianamente affida alla propria pagina Facebook i messaggi politici e la diffusione delle idee del proprio movimento». Diversamente, invece, ha deciso lo stesso Tribunale di Roma, con un'ordinanza più estesamente motivata[24], rigettando il ricorso presentato da un'altra associazione di estrema destra, Forza Nuova, cui pure era stata oscurata la pagina sul social network; in questo caso il giudice ha affermato il «dovere legale» di Facebook di rimuovere contenuti illeciti, non solo per violazione delle condizioni contrattuali, ma anche per violazione di un sistema normativo multilivello, in cui entrano in gioco il diritto internazionale, il diritto dell'Unione europea, quello della Convenzione europea dei diritti dell'uomo e il diritto nazionale, oltre che una «vasta giurisprudenza nazionale e sovranazionale», dettagliatamente citata nella pronuncia.

23 Tribunale di Roma, sez. impresa, 11 dicembre 2019 (R.G. 59264/2019).

24 Tribunale di Roma, sez. diritti della persona e immigrazione civile, 24 febbraio 2020 (R.G. 64894/2019).

E' recentissima, invece, la vicenda relativa alla chiusura del profilo social di Donald Trump: con una mossa senza precedenti, Facebook e Twitter hanno bloccato gli account del Presidente uscente USA dopo che costui aveva pubblicato false accuse sull'integrità delle elezioni, mentre i suoi sostenitori hanno assaltato il Campidoglio. Ne è derivata una polemica, che divide coloro che ritengono si sia trattato di una sostanziale censura, e coloro che invece sostengono che si sia trattato di un atto di responsabilità dovuto. In ogni caso la decisione di facebook ha portato all'attenzione la questione della necessità di regolamentare il potere dei *social* e il ruolo di controllo che essi esercitano sulle *fake news*, creando un precedente che affida alle piattaforme un potere editoriale di enorme portata: i *social*, come si è già ricordato, non sono enti editoriali, per cui non dovrebbero effettuare il controllo dei contenuti, né degli utenti; è accettabile che un imprenditore privato possa decidere chi può parlare alla gente? Non si può negare, infatti, che il riconoscimento di un simile potere ad un soggetto privato potrebbe, in futuro e in altre circostanze, innescare dilemmi ancora più complicati.

Queste vicende, che nascono dalla difficoltà del mondo dell'informazione ad adattarsi al modello economico imposto dai *social networks*, evidenziano ulteriormente la necessità di una regolamentazione della comunicazione politica sulle piattaforme, mettendo in luce la debolezza dell'approccio normativo fondato sull'autoregolamentazione dinanzi alla portata e alla rilevanza degli interessi costituzionali che devono essere tutelati. La dimensione costituzionale del problema è tale da imporre, piuttosto, un intervento legislativo vero e proprio; dovrebbe trattarsi però di un intervento normativo calibrato sulle caratteristiche tecniche dei *social* come strumento di costruzione del consenso, con la previsione, per quanto attiene alla comunicazione politica, ad esempio di divieti di profilazione degli utenti e di utilizzo di *bot/troll* e altri simili meccanismi di manipolazione del dibattito.

Problema diverso, benché strettamente interconnesso, è quello delle sanzioni applicabili per le *fake news* e dell'incidenza della disinformazione sulle scelte politiche dei cittadini. In questo caso si pone la necessità di interventi normativi più incisivi, la cui difficoltà si ascrive alla portata delle questioni coinvolte.

L'AGCom, nelle Linee guida sopra richiamate, ha invitato le piattaforme digitali ad effettuare un controllo sulle informazioni diffuse in rete. Anche in questo caso, si tratta solo di un invito, e non di un obbligo, che, peraltro sarebbe difficilmente configurabile: la notizia falsa per l'ordinamento italiano

non è in sé illecita (l'art. 656 c.p. proibisce la diffusione di notizie false esagerate o tendenziose, ma soltanto quando questa diffusione comporti un turbamento dell'ordine pubblico, che rappresenta un valore costituzionale volto ad assicurare la pacifica convivenza); ad esempio il negazionismo non può essere punito in sé, ma solo quando costituisca un'aggravante rispetto ad altri reati, come il reato di propaganda nazista o di incitamento all'odio. In casi del genere è possibile effettuare una segnalazione al gestore della piattaforma internet, che è tenuto a rimuovere contenuti lesivi di diritti altrui. D'altra parte, l'idea di affidare a soggetti pubblici o privati il compito di vagliare la veridicità delle notizie trasmesse in rete, e conseguentemente, di selezionarle decidendo sulla loro pubblicazione nelle piattaforme è incompatibile con i principi costituzionali connessi alla libertà di manifestazione del pensiero, e, a monte, la circoscrivibilità di tali misure ai rapporti di natura politica costituisce di per sé un problema di non facile soluzione.

Il tema, del resto, ha un raggio ancora più ampio, che potrebbe condurre ad affrontare la possibilità di utilizzare i meccanismi di funzionamento dei *social network* quali strumenti per costruire un ambiente deliberativo adeguato, incrementando le garanzie di pluralismo delle informazioni e delle opinioni. In altri termini, ci si può interrogare sulla possibilità di prevedere, di fronte a *hate speech* o di acclarate *fake news*, oltre alle sanzioni, anche veri e propri obblighi di controinformazione, oppure di ipotizzare rimedi come la rottura delle c.d. *filter bubbles* attraverso la promozione attiva dell'incontro con un pensiero divergente. Spingendosi ancora oltre si potrebbe prevedere che in campagna elettorale i soggetti politici, intesi anche come individui, siano chiamati a praticare la comunicazione politica mediante appositi profili, da sottoporre a regole apposite, con il presidio, temporaneo, dell'autorità indipendente. Ci si muoverebbe, ovviamente, su un terreno irto di difficoltà connesse alla limitazione di diritti fondamentali dei singoli, e tuttavia, il diritto costituzionale non può esimersi, data la rilevanza degli interessi in gioco, dal tentativo di affrontare una simile sfida.

4 CAMPAGNE ELETTORALI DIGITALI E PRINCÌPI COSTITUZIONALI: I RISCHI PER LA DEMOCRAZIA E LA NECESSITÀ DI UN INTERVENTO REGOLATORIO DEL LEGISLATORE

Le questioni coinvolte nella tematica considerata sono, come si è visto, rilevantissime e di difficile soluzione, tuttavia, gli ordinamenti contemporanei non possono rinunciare all'idea di configurare una responsabilità giuridica

dei *social* media, che faccia da contrappeso al ruolo svolto dalle piattaforme nella formazione del convincimento politico degli elettori.

Il momento della campagna elettorale è certamente quello in cui emergono in modo più evidente i rischi e le contraddizioni che caratterizzano la relazione tra la diffusione dei *social* media e i presupposti costituzionali della democrazia: il diritto di proprietà che l'impresa esercita sulla propria piattaforma, dal sito *web* ai diritti di proprietà intellettuale sul *software*, non può arrivare al punto di interferire sulla libertà di espressione, sul diritto all'informazione e sulla corretta formazione della rappresentanza politica.

Si è cercato di evidenziare la limitatezza dell'attuale approccio regolativo nell'ordinamento italiano: né le norme in materia di protezione dei dati personali, né quelle che disciplinano l'informazione e le campagne elettorali sembrano in grado di cogliere il problema nella sua interezza e, conseguentemente, forte è il rischio di zone d'ombra nella tutela dei diritti fondamentali.

Le dinamiche di interazione sui *social networks* presentano aspetti del tutto contrastanti con i principi costituzionali, poiché sono costruite sulla base di algoritmi che profilano e selezionano gli utenti secondo le loro preferenze. Questi meccanismi, applicati all'ambito della comunicazione politica, determinano la possibilità di favorire l'incontro tra opinioni analoghe, rafforzando i convincimenti precostituiti, ed escludendo qualsiasi possibilità di confronto tra diversi, se non in termini di scontro radicale (*confirmation bias*, polarizzazione delle opinioni). Inoltre, la personalizzazione della comunicazione consente ad uno stesso soggetto di indirizzare messaggi diversi a porzioni segmentate del proprio pubblico (*microtargeting*). Più in generale, le caratteristiche tecniche dei media digitali, che si avvalgono di profilazione, *troll*, messaggi automatici, *bot*, sono tali da offrire strumenti di persuasione artificiale e ingannevole. Per questo motivo le piattaforme sono estremamente funzionali all'insediamento alla diffusione di meccanismi formidabili di disinformazione.

Come già aveva evidenziato D. Zolo, sia pure in relazione al contesto mediatico precedente alla rivoluzione informatica, nell'ambito del paradigma neoclassico della democrazia pluralistica, il cittadino è ridotto a un consumatore politico (ZOLO, 1996), soggetto che si trova in una condizione deteriore rispetto al consumatore economico: infatti, mentre il consumatore economico gode di un certo controllo sui beni e servizi che acquista, il consumatore politico è privo della possibilità di controllare i

prodotti politici che gli vengono offerti, trovandosi nell'impossibilità di scegliere razionalmente quello che, sulla base di un'analisi costi-benefici, si riveli il più adatto alle sue preferenze.

La condizione tratteggiata da Zolo sembra ulteriormente accentuata nell'attuale sistema mediatico, in cui i *social* detengono una formidabile capacità di anticipare gli interessi e le preferenze degli utenti: in questo contesto emerge il rischio che gli attori politici utilizzino le piattaforme per implementare strategie idonee a manipolare la volontà degli elettori. L'esperienza recente evidenzia il fatto che i rischi connessi all'impatto dei *social* sull'informazione politica non sono solo teorici, se si pensa che è stato dimostrato che il trattamento dei dati personali e il *microtargeting* sui social hanno consentito alle forze politiche di mettere in atto strategie comunicative volte a disincentivare la partecipazione alle elezioni dei cittadini profilati come "potenzialmente avversi" (GREEN, ISSENBERG, 2017), o, al contrario, di spingere al voto gli elettori indecisi, o di diffondere messaggi di disinformazione per influenzare la volontà politica degli elettori.

Tutto questo non può non avere refluenze sui sistemi istituzionali, spingendo la trasformazione delle democrazie liberali verso sistemi in cui i cittadini, sottoposti alla sorveglianza pervasiva e al condizionamento da parte dei soggetti portatori di interessi economici e politici, avranno una capacità sempre più ridotta di formare e perseguire scelte che siano frutto di autonomia.

E' un'evoluzione che contraddice nettamente i princìpi della democrazia liberale pluralista, che si fonda sul riconoscimento dell'autonomia individuale, con riferimento ad un individuo capace di governare razionalmente le proprie scelte senza subire il condizionamento della sua volontà ad opera dei *social*. Se la democrazia è il potere in un popolo informato, come già riconosceva Toqueville nel diciannovesimo secolo, la società della disinformazione si avvia verso una pericolosa deriva: il contesto delle informazioni in cui gli utenti della rete sono immersi è tale da incidere sull'autodeterminazione politica dei soggetti, imprimendo trasformazioni profonde sui rapporti di potere, e, dunque, sulle istituzioni.

La comunicazione politica, e, segnatamente, le campagne elettorali, sono cruciali in questa degenerazione: le campagne digitali, in cui non vi è la parità di accesso e di parola da parte dei partecipanti, la presenza di una base informativa affidabile, l'inclusività di tutti i punti di vista rilevanti, non

consentono la realizzazione dei princìpi del pluralismo che dovrebbe essere l'essenza di ogni confronto democratico.

In un contesto in cui il flusso di informazioni attraverso la rete è gestito da parte di pochissime imprese transnazionali, la comunità giuridica è chiamata a compiere due esercizi di responsabilità: da un lato bisogna studiare di più i processi di cambiamento per non applicare a mondi nuovi parole e paradigmi del passato. Affinché il giurista non rimanga confinato nel deserto dell'astrattezza è necessario un approccio interdisciplinare, che faccia riferimento anche alle scienze computazionali per elaborare gli strumenti tecnici necessari ad interpretare, procedimentalizzare e risolvere i nuovi conflitti giuridici. Solo così è pensabile, infatti, tentare di affrontare la complessità delle relazioni tra l'evoluzione tecnologica e il diritto. In secondo luogo, occorre attribuire una nuova centralità al controllo del pluralismo nel nuovo contesto mediatico, per evitare che continuino a passare sotto silenzio fenomeni di *favor*, di conflitto di interesse e di *impar condicio*.

BIBLIOGRAFIA

AMPOFO L., COLLISTER S., O'LOUGHLIN B., CHADWICK A., *Text Mining and Social Media: When Quantitative Meets Qualitative, and Software Meets Humans*, in Halfpenny, Procter (eds.), *Innovations in Digital Research Methods*, London, Sage, 2015.

ANNANY M., *Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness*, in *Science, Technology and Human Values*, vol. 41, 2016, n. 1, p. 97.

AA.VV., *Social network, formazione del consenso e istituzioni politiche: quanto hanno influito i social sulle elezioni europee?*, in *www.gruppodipisa.it,* 24 giugno 2019.

AVVISATI M., *A.G.COM. e par condicio al tempo di internet*, in *Osservatorio sulle fonti*, 2014, n. 2, p. 10.

BECKETT L., *Trump Digital Director Says Facebook Helped Win the White House, www. theguardian.com/technology/2017/oct/08/trump-digital-director-brad-parscale-facebook advertising*, 2017; *How Facebook Ads Helped Elect Trump, www.cbsnews.com/news/ how-facebook-ads-helped-elect-trump*, 2017.

BENNET C.J., *Voter Databases, Microtargeting, and Data Protection Law: Can Political Parties Campaign in Europe as They Do in North America?*, in *International Data Privacy Law*, vol. 6, 2016, n. 4, p. 261.

BESSI A., FERRARA E., *Social Bots Distort the 2016 U.S. Presidential Election Online Discussion*, in *First Monday*, vol. 21, 2016.

BIN R., *Casa Pound vs. Facebook: un'ordinanza che farà discutere*, in *lacostituzione.info*, 15 dicembre 2019.

BOND R. M., FARISS C.J., JONES J.J. et al., *A 61-Million-Person Experiment in Social Influence and Political Mobilization*, in *Nature,* vol. 489, 2012, pp. 295-298.

CALISE M., MUSELLA F., *Il Principe digitale*, Roma-Bari 2019.

CARAVITA DI TORITTO B., *Social network, formazione del consenso, istituzioni politiche: quale regolamentazione possibile?*, in *Federalismi.it*, n. 2/2019.

CHESTER J., MONTGOMERY K., *The Role of Digital Marketing in Political Campaigns*, in *Internet Policy Review*, vol. 6, 2017.

COLONNA L., *A Taxonomy and Classification of Data Mining*, in *Science and Technology Law Review*, vol. 16, 2013, p. 309.

CUNIBERTI M., *Tecnologie digitali e libertà politiche*, in *Il Diritto dell'informazione e dell'informatica*, 2015, n. 2, p. 275.

DELANY C., *How Political Campaigns and Advocates Can Use Social Media Data*, *www.epolitics.com/2014/01/14/howpolitical-campaigns-and-advocates-can-use-social-media-data*.

DE MINICO G., *Pubblicità elettorale on line: regole o anarchia*, intervento al forum *Le sfide della democrazia digitali*, in *www.gruppodipisa.it*, n. 3, 2019, p. 231.

EPSTEIN R., ROBERTSON R., *The Search Engine Manipulation Effect (SEME) and Its Possible Impact on the Outcomes of Elections*, in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, 2015, n. 33.

FARO S., LETTIERI N. (eds.), *Law and Computational Social Science*, Napoli, 2013

FLORIDI L., *La quarta rivoluzione. Come l'infosfera sta trasformando il mondo*, Milano, 2017

FOSTER R., *News Plurality in a Digital World*, London, Reuters Institute for Journalism, 2012.

GOLIA A., BEHRING R., *Private (Transnational) Power without Authority. Online fascist propaganda and political participation in CasaPound v. Facebook*, in *Verfassungsblog*, 18 febbraio 2020.

GOODMAN E., LABO S., TAMBINI D., MOORE M., *The New Political Campaigning*, in *Media Policy Brief 19*, London, The London School of Economics and Political Science, 2017.

GRASSO G., *Social network, partiti politici e lotta per il potere*, in *MediaLaws*, n. 1, 2020.

GREEN J., ISSENBERG S., *Inside the Trump Bunker, With Days to Go, www. bloomberg. com/news/articles/2016-10-27/inside-the-trump-bunker-with-12-days-to-go*, 2017.

HERSH E.D., *Hacking the Electorate: How Campaigns Perceive Voters*, Cambridge, Cambridge University Press, 2015.

HOWARD P.N., KOLLANYI B., BRADSHAW S., NEUDERT L.M., *Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing States?*, *http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/09/Polarizing-Content-and SwingStates.pdf,* 2017.

KOHL U., *The Rise and Rise of Online Intermediaries in the Governance of the Internet and Beyond – Connectivity Intermediaries*, in *International Review of Law, Computers and Technology*, vol. 26, 2012, n. 2-3, p. 185; Id., *Google: The Rise and Rise of Online Intermediaries in the Governance of the Internet and Beyond* (Part 2), in *International Journal of Law and Information Technology*, vol. 21, 2013, n. 2, p. 187.

KRANZBERG M., *Technology and History: "Kranzberg's Laws"*, in *Technology and Culture*, 1986, Vol. 27, pp. 544-560.

KREISS D., *Prototype Politics: Technology-intensive Campaigning and the Data of Democracy*, Oxford, Oxford University Press, 2016.

KROLL J.A., HUEY J., BAROCAS S., *Accountable Algorithms*, in *University of Pennsylvania Law Review*, vol. 165, 2017, p. 633.

LAZER D., PENTLAND A., ADAMIC L. et al., *Computational Social Science*, in *Science*, vol. 323, 2009, n. 5915, p. 721.

LUMB D., *Why Scientists Are Upset About The Facebook Filter Bubble Study*, https://*www. fastcompany.com/3046111/why-scientists-are-upset-over-the-facebook-filter-bubble-study*, 2015.

MARTIN REYES J., *Social network, polarizzazione e democrazia: dall'entusiasmo al disincanto*, in E. Vitale – F. Cattaneo (a cura di), *Web e società democratica. Un matrimonio difficile*, Torino, 2018, p.18.

MELZI D'ERIL C., VIGEVANI G. E., *Facebook vs Casapound: un social network è davvero un servizio pubblico?*, in *Il Sole 24 Ore*, 15 dicembre 2019.

MICONI S., *Comunicazione e pubblicità istituzionale: classificazioni e regolamentazione*, in *Il Diritto dell'informazione e dell'informatica*, 2016, n. 6, p. 901.

MONTALDO R., *Le dinamiche della rappresentanza tra nuove tecnologie, populismo, e riforme costituzionali*, in *Quaderni costituzionali*, 4, 2019, 789.

NAPOLI P., *Digital Intermediaries and the Public Interest Standard in Algorithm Governance, http://blogs.lse.ac.uk/mediapolicyproject/2014/11/07/digital-intermediaries-andthe-public-interest-standard-in-algorithm-governance/,* 2014.

PARISER E., *The Filter Bubble: What the Internet Is Hiding from You*, New York, Penguin, 2011

PASQUALE F., *The Black Box Society: The Secret Algorithms That Control Money and Information*, Cambridge, Harvard University Press, 2015.

REED M., *Social Network Influence on Electoral Outcomes, https://ssrn.com/abstract= 2349273*, 2013.

RODOTÀ S., *Tecnopolitica. La democrazia e le nuove tecnologie della comunicazione*, Roma, 2004.

SCIORTINO A., *Fake news and infodemia at the time of Covid-19*, in *RDP Brasilia*, n. 17/2020, p. 35 e ss.

STAMOS A., *An Update on Information Operations on Facebook, https://newsroom.fb. com/news/2017/09/information-operations-update*, 6 settembre 2017.

SUNSTEIN C. R., *Infotopia: How Many Minds Produce Knowledge*, New York, Oxford University Press, 2006.

TALBOT D., *Facebook: The Real Presidential Swing State*, in *MIT Technology Review*, vol. 5, 2012.

TAURO A., *La comunicazione delle pubbliche amministrazioni in periodo elettorale*, Milano, 2016.

ULBRICHT L., M. VON GRAFENSTEIN M., *Big Data: Big Power Shifts?*, in *Internet Policy Review*, vol. 5, 2016, n. 1.

ZOLO D., *Il principato democratico*, Milano, 1996, p. 172.

ZOLO D., *The "Singapore Model": Democracy, Communication, and Globalization*, in K. Nash, A. Scott (eds.), *The Blackwell Companion to Political Sociology*, Oxford, 2001, p. 415.

ZOLO D., *Il tramonto della democrazia nell'era della globalizzazione, http://www.juragentium. org/topics/wlgo/it/tramonto.htm, 2010.*

ZUBOFF S., *The Secrets of Surveillance Capitalism*, in *Frankfurter Allgemeine Zeitung*, 5 giugno 2016.

ZUBOFF S., *Big Other: Surveillance Capitalism and the Prospects of an Information Civilization,* in *Journal of Information Technology*, vol. 30, 2015, n. 1, p. 75.

ZUIDERVEEN BORGESIUS F., TRILLING D., MÖLLER J. et al., *Should We Worry About Filter Bubbles?*, in *Internet Policy Review*, vol. 5, 2016, n. 1.

Sobre a autora:

Ornella Spataro | *E-mail:* ornella.spataro(at)unipa.it

Laurea in Giurisprudenza conseguita presso l'Università di Palermo il 27 aprile 1995 con la votazione di 110/110 e lode. Dottorato di ricerca in Diritto Comunitario e Diritto Interno. Fonti, Organizzazione, Attività, conseguito presso l'Università di Palermo (esame finale nel 2000). Borsa di studio denominata "Premio Giovanni Bonsignore", istituita con L.R. Sicilia del 16.05.1991, n. 27, conseguita nel 1999 a seguito di selezione per esami e titoli. Diploma in Public Management conseguito presso il CERISDI (Centro ricerche e studi direzionali) a seguito di programma annuale di studio e ricerca.

Artigo convidado.

Prolegômenos a uma Filosofia Algorítmica Futura Que Possa Apresentar-se Como Fundamento para um *Cyberdireito*

Prolegomena to a Future Algorithmic Philosophy That Could Present Itself as a Basis for a Cyberlaw Theory

MARIAH BROCHADO[1]

Faculdade de Direito da Universidade Federal de Minas Gerais (UFMG).

RESUMO: As concepções aqui apresentadas são o ponto de partida para se refletir sobre os fenômenos decorrentes do advento da revolução tecnológica e que passaram a ser recorrentes no cotidiano dos humanos, tais como o *big data,* a inteligência artificial, a rotina das redes sociais, as descobertas no campo da autonomia de máquinas, os desconfortos éticos surgidos destas e de outras experiências tão novas quanto insólitas, todas unidas por uma tessitura comum: a radical diferença do mundo analógico para o mundo digital. Certamente as filosofias tradicionais não se prepararam para enfrentar esse novo objeto de investigação sustentado por um mundo constituído por algoritmos. Os dilemas éticos surgidos exigem saídas jurídicas que as ciências jurídicas clássicas também não vislumbravam, visto que não conformadas à realidade virtual. Por essas razões, lançamos a provocação sobre a possibilidade de pensarmos numa filosofia algorítmica e num cyberdireito inspirados pelos desafios deste século.

PALAVRAS-CHAVE: Filosofia algorítmica; cyberdireito; algoritmização; inteligência maquínica; Direito e tecnologia.

ABSTRACT: The conceptions presented here are the starting point for reflecting on the phenomena resulting from the advent of the technological revolution and which have become recurrent in the daily lives of humans, such as Big data, artificial intelligence, the routine of social networks, discoveries in the field of machine autonomy, the ethical discomforts arising from these and other experiences as new as unusual, all united by a common texture: the radical difference from the analog world to the digital world. Certainly traditional philosophies were not prepared to face this new object of investigation supported by a world constituted by algorithms. The ethical dilemmas that emerged require legal solutions that classical legal sciences did not envision either, since they were

1 Orcid: https://orcid.org/0000-0001-5863-7360.

not conformed to virtual reality. For these reasons, we challenge the possibility of thinking about an Algorithmic Philosophy and a Cyberlaw inspired by the challenges of this century.

KEYWORDS: Algorithmic philosophy; cyberright; algorithmization; machinic intelligence; Law and technology.

1 INTRODUÇÃO E CONTEXTUALIZAÇÃO DA PROPOSTA

As discussões sobre os feitos da revolução tecnológica e da internet 4.0 têm girado em torno da parafernália que passou a fazer parte do nosso dia a dia, de como tirar proveito dela e como frear seus impactos sobre comportamentos, decisões e monopólio de agendas públicas e da vida dos consumidores em escala planetária. As indagações éticas são tratadas a reboque, restringidas, por um lado, a impressões do senso comum e, por outro, a previsões escatológicas inspiradas em obras ficcionais, aguardando que um grupo de pensadores independentes dos interesses do Vale do Silício e da linguagem *siliconense* venha assumir a tarefa de responder a dilemas éticos que não se colocavam há poucas décadas. A importância desta tarefa, objetivando enfrentar temas relacionados à existência humana datificada e conduzida por algoritmos, foi anunciada aos meus ouvidos pela primeira vez no ano de 1998 numa aula de Ética no antigo Instituto Santo Inácio de Loyola, quando a abordagem desse tipo de assunto no campo das ciências humanas era ainda incipiente no Brasil, tão incipiente que só nesta primeira quadra de século começa a ganhar musculatura. O tema surgiu acanhado como orbe das ciências exatas, sem chamar a atenção dos sistemas filosóficos do século passado, que seguiram alheios ao turbilhão de novas *formas de estar no mundo* que nos atingiriam rapidamente já nos anos 2000, e que despencariam sobre nossas vidas como uma tempestade na forma de dispositivos digitais inteligentes, de rastreabilidade do cotidiano promovida pela computação em tempo real, de plataformas parasitárias do "capitalismo de plataforma", do consumismo informacional desenfreado (pode-se padecer de falta de comida hoje, mas não de conteúdo [informacional] – Morozov, 2018, p. 49). Essas novas experiências são conquistas radicais iniciadas em meados do século passado e *hiperintensificadas* neste, e vêm inundando exponencialmente nossa vida de *technoartefatos* sem os quais não conse-

guimos mais viver (orelhão público é bibelô do século passado), alagando nosso cotidiano de cliques e *touches* que consomem nosso tempo, nossa energia e nosso dinheiro na tentativa vã de aprender a lidar com cada novidade que deveria tornar nossa vida mais confortável, mas que nos envolve num cipoal confuso e inadministrável do qual não conseguimos nos desvencilhar, pois não é mais possível fazer diferente: a *era analógica* vive seu derradeiro suspiro.

Lá em 1998, o avanço tecnológico ainda não havia atingido seu apogeu, e o debate sobre seus feitos era restrito a circuitos acadêmicos e laboratoriais, sequer constava do vocabulário popular, ao qual, ao lado do "eu tenho uma Brastemp", somavam-se timidamente rudimentares expressões como "consegui comprar um Motorola MicroTAC *com flip*" ou "substituí meu tijolão Motorola por um *telefone em barra* Nokia 6160". Eis que, apenas duas décadas depois, deparamo-nos com a seguinte campanha de *marketing*:

> O Boticário lança 1ºs perfumes feitos com ajuda de inteligência artificial. Depois de mais de dois anos de desenvolvimento, o grupo paranaense Boticário, maior varejista de cosméticos do país, lança os primeiros perfumes *do mundo* feitos com ajuda de inteligência artificial (AI, na sigla em inglês). Antecipada ao *site* Exame, a novidade chega em duas versões – "On you" (Em Você) e "On me" (Em Mim) – e integra o portfólio da marca Egeo, voltada para o público mais jovem.[2]

Voltando à aula de 1998 mencionada, a sacudida quanto à urgência do enfrentamento de temas afetos ao desenvolvimento galopante das tecnociências veio de uma incisiva e preocupada fala do filósofo mineiro Henrique Cláudio de Lima Vaz, cuja síntese ficou registrada em sua obra *Escritos de Filosofia V: introdução à ética filosófica II*, publicada em 2000. A advertência do filósofo é fundamental no desenrolar deste trabalho, daí a necessidade de citá-lo *ipsis litteris* já na introdução:

> Entre as causas que justificam a denominação de "século da Ética" proposta por alguém para o século que vai começar (outros preferem antecipar sua face cultural como "século da Biologia") estão, sem dúvida, de um lado, o vertiginoso crescimento das tecnociências, em particular da biotecnologia, e, de outro, a não menos rápida e abrangente dissolução do tecido social

2 Disponível em: https://www.google.com.br/amp/s/exame.com/marketing/o-boticario-lanca-1os-perfumes-feitos-com-ajuda-de-inteligencia-artificial/amp/. Acesso em: 20 jun. 2021.

> tradicional e a sua substituição por novas e inéditas formas de convivência humana e de organização da sociedade [...] Estamos assistindo a um processo de *mutação* (para usar uma metáfora biológica) muito mais profundo do que qualquer *mudança* e que assinala, provavelmente, o surgimento de uma nova civilização. Que *sentido* terá a vida humana nessa nova civilização? Que *valores* a guiarão? Que *fins* o ser humano poderá entrever para o seu caminho histórico? Essas são interrogações de natureza *ética* que solicitam com urgência alguma forma de resposta ao acelerado processo que nos arrasta em direção a terras ignotas de cultura e civilização. (Vaz, 2000, p. 240)

Em 2019, ano em que O Boticário lança a primeira linha de perfumes "feitos com inteligência artificial", as advertências de Lima Vaz ecoam como um ditirambo potente, conclamando nossa responsabilidade sobre a urgência de uma profunda reflexão sobre os rumos da *tecnocência*. Certamente naquele final da década de noventa, o filósofo jesuíta, com toda a sua monumental erudição e capacidade de prospecção, não acreditaria num anúncio dessa natureza. Como assim? Perfumes que trazem o nome de "Em você" e "Em mim", algo demasiado humano, destinado a humanos e há séculos pensado, elaborado, experimentado e sentido *exclusivamente* por humanos, agora é propagandeado como peculiarmente melhor elaborado por ser feito por *máquina* supostamente mais *inteligente* que humanos? Onde estão os perfumistas, aqueles notáveis humanos que se destacavam por sua sensibilidade olfativa e capacidade de realizar as melhores combinações de odores em diapasão de notas olfativas de flores, frutos, madeiras, folhas? Teremos também a inteligência artificial costumizando os melhores vinhos ainda não inventados em séculos de história da enologia?

Aqui começa o nosso desafio: fragrâncias e sabores são *qualia* e é impossível conceituar essas experiências humanas com precisão. O que nós humanos fazemos é compartilhar estas impressões indefiníveis por aproximação de percepções *singulares* e que são em nós provocadas por incontáveis afecções pelo contato com o mundo físico por meio dos nossos sentidos. O nosso corpo é o veículo dessas sensações e através dele é que as compartilhamos, sempre nos esforçando para elaborar mentalmente noções que as representem na forma de comunicação. É bem provável que Lima Vaz tenha pressentido que o porvir (que ele não testemunharia) era gigantescamente inovador ao fazer esta predição num sóbrio e aflitivo exercício intelectual nada futurológico, como podemos constatar nas linhas transcritas, mas nem de longe sua intuição limitada no tempo-espaço possibilitou-lhe a predição de detalhes do que aconteceria com a biotecnologia e a tecnologia

da informação e que nos arrastariam "em direção a terras ignotas" em tão poucos anos. Ele jamais suporia que nestas terras ignotas *artificiais* receberíamos de rebarba os desagradáveis atendimentos por *chatbots* toda vez que buscamos um produto na internet, e os (mais desagradáveis ainda) presidenciáveis eleitos à custa da indústria de *fake news*. Mais: ele não preditaria que a presença dos humanos em várias tarefas banais, como as de ouvir, dialogar e solucionar problemas com outros humanos, seria dispensada e substituída por *assistentes virtuais*; e, pior, num segundo momento, acoplada às novas *formas maquínicas* de tratativa muito mais eficientes. É assim, por exemplo, que nos é ofertado outro produto híbrido *techno-humano* quando recebemos pelo Facebook um anúncio do tipo: "Globalbot é uma solução de atendimento a clientes com inteligência artificial que combina *chatbots* (assistentes virtuais) com atendimento humano em diversos canais"[3]. Temos que admitir que recursos humanos tornaram-se obsoletos e dispensáveis.

As leituras e as reflexões que tenho empreendido nos últimos anos são, no fundo, uma tentativa de resposta à dramática *provocação ética* lançada pelo mestre Padre Vaz. Cônscia da impotência face ao medo quanto à impossibilidade mesmo de alguma derradeira resposta, procuro identificar, neste artigo, alguns questionamentos que se impõem às ciências humanas na encruzilhada civilizacional que vivemos hoje, e diante da qual o drama que martiriza o filósofo que se propõe a refletir sobre ética e direito pode ser resumido em quatro indagações básicas, em torno das quais todo o resto orbita: o que distingue máquinas pensantes (e aprendizes) de seres ditos *conscientes*? A *senciência*, afinal, é um emaranhado de *algoritmos bioquímicos* forjados por milhares de anos em atividades neurais que caracterizam, na escala evolutiva, o *homo sapiens*, o que ainda será decifrado e reproduzido com muito mais rigor e eficiência pelos feitos da biotecnologia? A consciência humana é atividade equivalente a conexões *regulatórias maquínicas*? Qual será o impacto destas descobertas e das novas concepções de mundo na seara jurídica, visto que o Direito em todo o seu percurso histórico até aqui constitui-se e equilibrou-se sobre o primado do intelecto, da vontade e da *liberdade*, faculdades tidas como *exclusivamente* humanas?

Feitas essas considerações de fundo, insta assinalar que o objetivo central deste artigo é lançar as bases sobre as quais podemos pensar numa

3 Disponível em: https://globalbot.com.br/?%7Bipurl%7D?keyword=sistema%20chatbot&gclid=CjwKCAjwrPCGBhALEiwAUI9X0zTIe4pGTJ5cnU4OvCZw4kfJeFIyCWIEaNjQkKRT0rGexzJ3UDHHSBoCjaAQAvD_BwE. Acesso em: 25 jun. 2021.

Filosofia Algorítmica e numa Ciência *Cyber* do Direito voltadas para as experiências e demandas do século que se inicia e que traz do século passado uma dívida incontornável: enfrentar as consequências da revolução tecnológica como pauta urgente no que se refere à mudança comportamental da humanidade e que ganhou proporções tão agigantadas ao ponto de forjar um novo tipo de vivência do *homo sapien*s, cujo cérebro, seu diferencial como sapiente na escala evolutiva dos terráqueos mamíferos, levou-o a sua própria superação pela criação de seres artificiais mais sapientes que ele mesmo. Esta é a razão mais significativa que move a narrativa que se desdobra neste trabalho, por intermédio de algumas chaves hermenêuticas que nos introduzirão com algum rigor no enfrentamento das questões aqui lançadas.

2 SOBRE FILOSOFIA E DIREITO

A cultura ocidental sempre se fiou na convicção de que a Filosofia é uma reflexão de *terceiro grau* sobre a ciência do seu tempo. Essa afirmação remonta às origens do pensamento científico, reconhecendo dois outros saberes que precedem o conhecimento filosófico: o conhecimento técnico, que integra o rol de saberes necessários para estar no mundo, e que, com menos rigor, aponta para um fazer (*techné*), um operar sobre a realidade com o objetivo de dominá-la sem grande aprofundamento e precisão; e o conhecimento científico (*epistéme*), aquele que é definido pelo rigorismo metódico na busca por padrões universais de explicação dos dados da realidade e que torna possível o aprofundamento nas especificidades que nos passam despercebidas nas aparições fenomênicas, possibilitando, portanto, maior controle sobre a realidade e fruição dela. Este conhecimento, dito científico, pretende ser tanto mais rigoroso quanto mais se aprofunda nas diversas áreas que se consolidaram desde o *enciclopedismo iluminista*, o qual intentou dissecar toda a realidade a partir dos estatutos epistemológicos das várias ciências, e que foram catalogadas, segundo a proposta positivista, em três grandes áreas: a das *ciências naturais*, a das *ciências hipotético-dedutivas* (lógico-matemáticas) e a das *ciências humanas* (Vaihinger, 2013; Machado, 1979 *apud* Kelsen, 1979).

Sendo nosso fio condutor transversal neste trabalho a Filosofia *do* Direito, assumimos a tradicional colocação deste saber dentre as ciências humanas, as quais se diferem das naturais pelo método de abordagem, a saber: i) as *naturais* partem da possibilidade de constatação de ligações entre fatos, deduzindo-se que deles derivam os fenômenos estudados; ii) as

humanas, às quais, para além das explicações próprias das ciências naturais, soma-se o ato de *compreender*, o que implica assumir que o cientista procura reproduzir intuitivamente o *sentido* dos fenômenos, atribuindo-lhe *valorações* (Ferraz Júnior, 1980, p. 11). Todas as ciências que seguem esse segundo padrão foram consideradas, sob a perspectiva metodológica, *ciências hermenêuticas* ou ciências *da compreensão*, diversas das ciências naturais, que são consideradas ciências *da explicação*. Importa posicionar a Hermenêutica neste tópico que introduz diferenças entre formas de conhecimento e dentre elas situa a Filosofia e o Direito porque a Cybernética a ela se assemelha no percurso epistemológico ocidental, o que, como veremos mais adiante, é uma informação *sine qua non* para traçarmos a proposta prolegômena à Filosofia Algorítmica e ao Cyberdireito, tal como anunciado no título desse artigo.

Voltando à catalogação triádica das ciências, ela é aceita sem reservas pela comunidade científica, não obstante estar impregnada de pré-compreensões legadas por empreitadas cientificizantes da experiência *cognitiva* humana, a exemplo do *ficcionalismo* de Hans Vaihinger, que aponta o artificialismo arbitrário destas categorizações do real. Vaihinger é um dos mais reconhecidos legatários de Kant, foi fundador dos *Annalen der Philosophie. Mit besonderer Rücksicht auf die Probleme der Als Ob Betrachtung* em 1919, e da *Kant-Gesellschaft*, em 1904. Ele se destacou como epistemólogo ao apresentar sua *Teoria do als ob* ("como se"), segundo a qual todas as ciências manejam *ficções* necessárias para justificar e unificar a pretensão das várias ciências em buscar os fundamentos últimos da realidade. Daí ele nomear sua tese de "*als ob*", no sentido de que as ciências dispõem de ficções sobre o real, tomando-as "como se fossem" verdadeiras, sabendo que são artifícios retóricos não sujeitos à verificação. São esquemas que visam a dar unidade a um sistema de observações e conclusões, mas eles mesmos são tomados arbitrariamente, o que, no discurso filosófico, conhecemos ou como axiomas ou como aporias.

Esse sistema ficcional garante pontos de partida seguros (*topois*) para as várias ciências, e a distribuição em três grandes áreas é uma manifestação desta necessidade, o que tem sido acolhido para definir disciplinas escolares, conteúdos programáticos dos cursos superiores, todas as divisões acadêmicas. Dessa natureza são as construções do conhecimento do tipo: a matéria se concebe "como se" fosse composta de átomos, o *eu* "como se" fosse uma substância, e tantas outras que permeiam as hipóteses e axiomas científicos calcados em ficções do intelecto (Vaihinger, 2013). Ocorre que,

tentando ir além das experiências imediatas da inteligência humana, o conhecimento filosófico visa a buscar as razões que permeiam o próprio *processo de conhecimento*, perquirindo categorias mais gerais que explicam a forma de *estar no mundo* do animal dotado de *racionalidade*, que é o *ser humano*. Tais afirmações parecem triviais se considerarmos a tradição filosófica ocidental, mas, no contexto deste artigo, releva registrar que a *condição de humano*, ou a dita *natureza humana*, ou a humanidade em nós, é o mote que desde sempre inspirou arquitetar uma filosofia e erguer entre suas divisões uma monumental plêiade de saberes, empreitada que vem desde a Grécia antiga apontando possibilidades de compreensão do mundo a partir da *situação do sujeito* (humano) neste mundo, seja ele o mundo físico, seja ele o mundo cultural.

Pensando em termos de situação das reflexões ofertadas pelas *ciências humanas* segundo aquela catalogação tradicional, precisamos somar a elas um esforço hermenêutico diferenciado na tradicional discussão sobre as relações entre Filosofia e Direito, aqui pensadas pela chave hermenêutica do advento da Cybernética e suas decorrências, dentre elas os processos de maquinização e algoritmização da realidade. Entendemos que a Cybernética não se enquadra exatamente nas tipologias epistemológicas tradicionais, o que nos leva a crer que ela se situa nas fronteiras de uma *nova forma de pensar* ciência (e de praticá-la) por ter surgido como a tentativa mais original de alguns cientistas ao apresentá-la como uma *superciência* unificadora de todas as demais, numa posição similar à qual ocupou a Hermenêutica enquanto unificadora das ciências culturais ou do espírito-*Geisteswissenschaften* (Dilthey, 1922). Em termos de estatuto epistemológico, a Hermenêutica surgiu como a *metaciência* que pretendeu *desregionalizar* os métodos de compreensão praticados por todas as ciências humanas e que poderiam ser unificados pela *Ciência ou Teoria da Interpretação* (Hermenêutica – grafado com maiúscula) (Schleiermacher, 1977; Ricoeur, 1969; Gadamer, 1994). Foi num contexto similar que a Cybernética se formou como ciência, carregando a ambiciosa pretensão de explicar *toda* a realidade, independente de divisões epistemológicas, tendo sido apresentada por seus artífices como a grande ciência que explicaria o mundo

> como um conjunto de sistemas de *feedback*, permitindo o controle racional de corpos, máquinas, fábricas, comunidades e praticamente qualquer outra coisa. A cibernética prometia reduzir problemas "confusos", em campos tais como a economia, a política e talvez a moral, à condição de simples tarefas de engenharia: uma coisa que se poderia resolver com lápis e papel ou,

> na pior das hipóteses, com um dos supercomputadores do MIT. (Haraway; Kunzru; Tadeu, 2009, p. 125)

Para seguirmos com a discussão sobre a posição da Filosofia e do Direito nesta abordagem, precisamos dar mais um passo para situá-los respectivamente como estatuto teórico *conduzido a algoritmos* e estatuto dogmático *conduzido ao controle* no sentido *cybernético*, objeto do próximo tópico. Trata-se de uma empreitada pouco louvável se entendida como semântica classificatória, mas ganha relevância pragmática se concordarmos que a Filosofia *do* Direito contemporânea tem por missão desvelar novas formas comportamentais talhadas num mundo *cybernetizado*, maquinizado, algoritmizado, e que trazem consigo a demanda por reconhecimento na qualidade de direitos fundamentais e não necessariamente humanos.

3 SOBRE CYBERNÉTICA COMO CIÊNCIA DE FRONTEIRA

Comecemos por justificar que a adoção do prefixo *cyber* grafado com *y* é proposital e nada tem a ver com anglicismo. O prefixo é hoje empregado de forma confusa e inconsistente, razão pela qual deixaremos anotada uma síntese da nossa opção. O prefixo decorre do grego *kyber*, cuja origem etimológica remetia à condução de embarcações, expressando também comando, governo; assim, rigorosamente o *cybernético* é o *timoneiro*, aquele que detém as rédeas do funcionamento, o controle da embarcação. O timoneiro troiano Palinurus é um personagem da mitologia romana encontrado no poema épico *Eneida*, do poeta romano Virgílio, datado do século I a.C. Ele era um experiente navegador que tinha por missão conduzir o navio do troiano Enéas, personagem principal da obra. No épico, quando os troianos deixam Cartago, Palinurus faz a *predição* de que uma tempestade se anuncia e de que Enéas não deve navegar até a Itália para participar dos jogos fúnebres em homenagem ao seu pai, Anquites. Para conduzir a embarcação com precisão, Palinurus aproxima-se das rochas, "obtém *informação visual* sobre a posição do barco e ajusta o curso de acordo com essa informação. Esse não é um evento singular, mas um *fluxo* constante de informação" que torna o timoneiro parte integrante "de um *circuito de feedback*, seu cérebro recebe um input do ambiente, que *informa* a velocidade do vento, o tempo e a corrente e envia, então, sinais para que seus braços possam conduzir seu barco para longe do perigo" (Haraway; Kunzru; Tadeu, 2009, p. 124).

É inspirada nesta conotação que a palavra "cybernética" entrou para a teoria do conhecimento ocidental. Remontando à filologia do termo, foi An-

dré Marie Ampère que criou a palavra *Cybernétique* em 1834, mencionando-a em sua obra *Essai sur la philosophie des sciences ou exposition analytique d'une classification naturelle de toutes les connaissances humaines*. Ele a adotou com o propósito de designar a inauguração de um novo rol de saberes que se ocupam do *estudo do pensamento* e dos meios pelos quais os homens, através da *comunicação*, vivem e se *governam*, integrando o que o filósofo e matemático chamou de *ciências noológicas* (Ampère, 1834). O termo entrou de vez para a história da ciência com o uso por Norbert Wiener, em *Cybernetics: or control and communication in the animal and the machine*, para indicar uma área de conhecimento voltada para o estudo do *controle* e da *comunicação* em animais e máquinas (Wiener, 1948).

Precisamente neste sentido, como ciência que estuda *sistemas regulatórios via comunicação* entre humanos, é que resgatamos aqui o conceito de Cybernética e mantivemos o prefixo grafado com *y*, tal como no original grego. Enquanto disciplina científica, a Cybernética é considerada uma "ciência de encruzilhada" integrada por um "conjunto de ciências limítrofes", tal como a Ontologia no âmbito filosófico, englobando estudos de lógica, matemática, neurofisiologia, engenharia, linguagem, entropia etc. Seus recortes assumem a tarefa de estudar *sistemas de controle e autorregulação*, como o de retroação (*feedback*), orgânicos e mecânicos, englobando os sistemas denominados *comportamentos com propósito*, e paralelismos entre sistemas de autocontrole e autorregulamentação, denominados *servomecanismos* (Mora, 2000, p. 453).

Daí ser imprescindível afirmar a importância do legado desta *ciência de fronteira* que impacta diretamente o Direito com as concepções de *autorregulação* e *regulação*, fundamentais para se compreender o funcionamento da *algoritmização* da vida humana e sua *anonimização* algorítmica. Falemos agora sobre regulação.

Os conceitos de regulação e autorregulação são trazidos para este trabalho no sentido proposto por Georges Canguilhem, que se inspirou nos clássicos escritos de Immanuel Kant. Na obra *Crítica da faculdade do juízo* (*Kritik der Urteilskraft*, 1789), Kant trouxe as definições de *organização* e *organismo* biológico como entidades dotadas de *reciprocidade interdependente* (Kant, 1993). Segundo o filósofo, são a reciprocidade de *partes* e a interdependência de *processos* que possibilitam a existência de sistemas fechados que constituem toda a realidade (Kant, 1993). O corpo humano é um sistema fechado que estabelece trocas com o meio ambiente. Se ele

passar para a condição de dilacerado ou putrefato, perde a característica de sistema fechado nesta primeira constituição e passa a integrar outro sistema, ao qual suas partes aderem.

Relacionando *organismo*, *máquina* e *sociedade*, Canguilhem adota tais bases à configuração de *organismos maquínimos*, o que supõe *regulação* e *autorregulação*. Esta possibilita a *totalidade* e o *fechamento* do sistema na forma de integração de subprocessos em uma estrutura causal; aquela, de maneira mais ampla, possibilita não só a organização, mas também a *manutenção* de sistemas organizados, incluindo aí suas relações internas e as *influências perturbadoras* a que ele está sujeito no *meio* em que se situa (Canguilhem, 2009; Toepfer, 2006):

> *nicht das Bestehen eines organisierten Systems, sondern dessen Erhaltung […] das Thema des Regulationskonzeptes (bildet). Die in der Biologie seit langem in einer stabilen Terminologie beschriebenen Phänomene der Ernährung, des Schutzes sowie der Koordination und Integration der Prozesse lassen sich zusammenfassend als Regulationen verstehen. Die Regulation betrifft also nicht nur die internen Bezüge eines Systems, sondern sie handelt wesentlich von der Beziehung eines organisierten Systems zu seiner Umwelt, insbesondere zu den von ihr ausgehenden störenden Einflüssen.* (Toepfer, 2006, p. 167)

De acordo com Canguilhem, a concepção kantiana de *organização* não foi suficiente para *separar* entidades orgânicas e inorgânicas, o que só foi possível posteriormente, com a inauguração da Cybernética (Canguilhem, 2009), quando esta nova ciência ganha espaço e passa a tratar das *entidades inorgânicas organizadas, reguladas* (estabilizadas) e em *evolução* (variação). A situação dos seres humanos no mundo, no plano natural e cultural, configura experiência dessa natureza, pois pressupõe *interdependência* e *reciprocidade*. Em termos cybernéticos, esta troca recíproca e interdependente constante se realiza pela comunicação (essencialmente autorregulada e regulada), sendo ela o meio pelo qual nos *conectamos*, seja manejando entidades físicas (tangíveis), seja manejando entidades abstratas (intangíveis) e *desterritorializadas* (Lèvy, 2018, p. 49). Nesse sentido, relações humanas são essencialmente *cyber-relações*. E a forma de manejo das relações hoje é a tecnologia.

A genialidade de Weiner foi perceber que a observação da realidade apresentada por Kant somada ao modelo desenvolvido por Ampére seria aplicável a qualquer problema que implicasse a administração de sistemas.

Sua ambição era de que os cientistas aplicassem os princípios da Cybernética a todas as investigações, já que

> pegar uma bola, guiar um míssil, administrar uma empresa, fazer o sangue circular em um corpo – tudo lhe parecia depender da transmissão de "informação", um conceito sugerido por Claude Shannonn, dos Laboratórios Bell, em sua obra fundadora sobre teoria da informação. Mais especificamente, esses processos pareciam depender daquilo que os engenheiros tinham começado a chamar de "*feedback*" [...] O sonho de Wiener, de uma ciência universal da comunicação e do controle, apagou-se com o correr dos anos. A cibernética deu origem a novas áreas como ciências cognitivas e estimulou pesquisas valiosas em numerosos outros campos. Mas quase ninguém, hoje, se auto intitula um "ciberneticista". (Haraway; Kunzru; Tadeu, 2009, p. 125)

Ainda que a empreitada de Weiner tenha sido esquecida, ela é convincente e atual, e, pelos desdobramentos no campo tecnológico, sua tese só se confirmou; e hoje precisamos resgatá-la. Especula-se que o grande projeto de Wiener acabou se tornando vítima do modismo científico e que muitos recursos foram desperdiçados em suas pesquisas sobre inteligência artificial sem retorno expressivo. Por outro lado, alguns acreditam que a concepção de Cybernética foi desconsiderada por uma falha original: "Os mecanismos básicos de controle e comunicação nas máquinas são significativamente diferentes daqueles que existem nos animais e nenhum deles se assemelha aos mecanismos de controle e comunicação existentes na sociedade" (Haraway; Kunzru; Tadeu, 2009, p. 125).

4 CYBERNÉTICA E TECNOLOGIA: IMPRECISÕES TERMINOLÓGICAS E CONFUSÕES CONCEITUAIS

Todas as formas de organização da realidade que manipula e emprega aparatos tecnológicos e realidades virtuais são desdobramentos de uma mesma ciência, e que tem sido esquecida como a ciência que deveria inspirar a filosofia deste século. Rigorosamente, referir-se a uma *Ciência da Computação* é apontar para um tipo de conhecimento adstrito a descrever, desenvolver e otimizar aparatos cybernéticos.

No vocabulário usual, tudo o que diz respeito a tecnociências é dito "tecnologia", mas as tecnociências nada mais são que desdobramentos da Cybernética, e a expressão *Ciência da Computação* não revela que ciência é esta, mas apenas afirma que ela se dedica ao sistema computacional, o qual é técnica e tecnologia, como veremos nas definições a seguir. Uma ciência dedicada a compreender e dominar um sistema de funcionamento

nada mais é que um estatuto *poiético* trazido como enfoque epistemológico das atividades que *fazem o sistema funcionar, produzir*. A atividade racional destinada ao *fazer* não pode ser confundida com a atividade racional destinada a *teorizar sobre* este fazer. Estas duas dimensões da inteligência humana fazem parte do repertório vocabular e conceitual da Filosofia, a qual distingue três dimensões da racionalidade humana: a razão poiética (voltada para o *produzir* coisas, dita razão fabricadora), a razão prática (a forma da razão na qual se exprimem os fins morais do *agir* humano e suas normas) e a razão teórica (ou razão *demonstrativa*, destinada a entender os fenômenos e dominá-los, formando as *ciências*).

> Na formação das capacidades cognoscitivas do indivíduo e na história dos grupos humanos, a *razão prática* antecede a *razão teórica* e é, sem dúvida, equioriginária com a *razão poiética* ou fabricadora. *Fazer* e *agir* são as duas primeiras atividades humanas conduzidas pela razão e que se manifestam simultaneamente na história das sociedades e dos indivíduos [...] [sendo a razão prática propriamente o que nos torna humanos], visto que "a pré-compreensão das regras primitivas do *fazer* que caracterizam o *homo habilis* admite analogias com as habilidades técnicas do animal, ao passo que o *agir* é atributo exclusivo do *homo sapien*. Nele está inscrita, mesmo em suas formas mais rudimentares, a pré-compreensão das razões normativas do agir". (Vaz, 2000, p. 26, 28)

As Ciência da Computação e a Ciência da Informação são partes da Cybernética, pois esta é verdadeiramente a ciência que engloba todas aquelas que pesquisam os sistemas tecnológicos, desde sua estrutura física e datificada até as consequências pragmáticas deles decorrentes. Não podemos, neste texto sintético, discutir a dispersão semântica do termo tecnologia, seus empregos em sentido técnico e vulgar. Mas deixemos claro que tecnologia é um modo de *artificializar a realidade*; nesse sentido, toda ciência, enquanto recorta um aspecto do real para dissecá-lo, é também tecnologia. Como salienta Evandro Agazzi:

> *ya en su comienzo la ciencia moderna se revela estructuralmente conectada a la tecnología, pues, en primer lugar, es necesario inventar, construir un instrumento para "observar" la naturaleza; en segundo jugar, el "experimento" científico consiste en la realización de una situación artificial, precisamente porque sólo dentro de una situación artificial se podrá poner a la vista lo que nunca se aprecia en una observación natural. Así, la ciencia experimental es una ciencia que ya, en su acta de nacimiento, lleva escrita la tecnología en sus raíces*. (Agazzi, 1998, p. 25)

O conceito de técnica se difere de tecnologia por uma distinção já feita entre os gregos. Para estes, a *téchne* era um conjunto conhecimentos eficazes que atuam concretamente sobre a realidade, mas não apenas como um amontoado de *formas de fazer* que se repetem porque dão certo, e sim por se saber que dão certo porque se sabe das razões por que dão certo. É saber por que determinadas práticas concretas são eficazes e, a partir deste saber, projetar outras sem a necessidade de experimentá-las previamente; tecnologia é, portanto, a teorização, a *cientifização da técnica*.

> *Esto muestra que en la civilización griega existía algo más que la mera acumulación, a veces simplemente casual, de experiencias que se transmitían de una generación a otra, reconociéndose que hay quienes, además de saber de la existencia de ciertos procedimientos eficaces, saben por qué lo son. Esta idea griega es la que ha quedado en ciertas expresiones, por ejemplo, cuando se dice que la "medicina es un arte", considerándose a ésta como un conjunto de prácticas eficaces que se apoyan en un conocimiento que justifica estas prácticas. Aquí se encuentra un preludio de la noción de tecnología. Cuando aparece el sufijo "logía" se quiere indicar la existencia de una cierta doctrina elaborada, una "teoría" acerca del asunto en cuestión (como cuando se habla de geo-logía, teo-logía, papiro-logía, antropo-logía, etc.). Así, en lugar de hablar de técnica hablamos de tecnología, añadiéndose algo más a la pura y simple técnica. La tecnología puede entenderse como aquello que acontece en el interior de la trayectoria de la técnica cuando surge, dentro de la civilización occidental, un conjunto de conocimientos "teóricos" que permiten explicar o dar razón de lo que es eficaz en concreto.* (Agazzi, 1988, p. 20)

Mas, ao discutirmos as *experiências humanas* sujeitas a estes sistemas, todos eles ligados por um objetivo comum – o *controle via comunicação* –, estamos falando de Cybernética. Esta a razão por insistirmos que ela deve ser posta no lugar que deve historicamente ocupar. Tal como a Matemática, a Física e a Biologia inspiraram as filosofias respectivamente de Platão, Kant e Hegel, e todo sistema filosófico se constitui sobre a ciência de seu tempo, é legítimo considerar a Cybernética a *ciência de fronteira* do nosso tempo e, portanto, responsável por pautar as reflexões filosóficas do tempo presente. E falar em ciência atualmente é também pensar em ciência da técnica, ou tecnologia, o que significa dizer que filosofar hoje é conhecer as ciências tradicionais e a tecnologia. Com Agazzi,

> *io ho sempre concepito la filosofia come uno sforzo di trovare risposte ai problemi fondamentali dell'esistenza umana, situata nel suo contesto storico-culturale. Constatando che la realtà contemporanea è permeata di scienza*

> *e tecnologia, mi apparve inevitabile che un'adeguata coscienza di ciò che significa tale presenza – vale a dire una profonda comprensione filosofica della scienza e tecnologia attuali – fosse la condizione indispensabile per risolvere i problemi esistenziali del mondo d'oggi*. (Agazzi, 2012)

Nossa primeira conclusão parcial é: não é possível enquadrar *conhecimento científico cyber* dentre os padrões epistemológicos lançados no tópico 2: ela não cabe nos modelos das ciências naturais, nem no das lógico-dedutivas e nem no das sociais, porque a realidade sobre a qual ela opera não é a natureza física (*phisys*), nem as atividades raciocinantes do intelecto humano (ainda que ciências lógicas se apliquem à Cybernética) e nem as sociedades humanas tal como foram concebidas pelos estatutos das ciências sociais. E não encontrar um lugar estratégico para a Cybernética na enciclopédia científica atual é um problema de fronteira que impacta a análise dos fenômenos ditos vulgar e inconsistentemente "tecnológicos". Voltaremos a essa questão.

5 SOBRE MAQUINIZAÇÃO, ALGORITMIZAÇÃO E ANONIMIZAÇÃO

O rastro científico que nos foi legado em mais de vinte e cinco séculos teve por propósito explicar, facilitar e evoluir as *experiências humanas*, e hoje essa evolução nos levou a uma encruzilhada existencial sobre *a reificação* humana que, de certa forma, *desumaniza o humano* face a suas próprias conquistas em ciência e tecnologia, o que vem ocorrendo como relação *homem-máquina*, ao ponto de nos referirmos, de forma banalizada, a uma dupla fenomenologia do espírito do século XXI: a *maquinização* do humano e a *humanização* da máquina, duas grandes referências para as fronteiras da ciência hoje.

O primeiro legado indiscutível da Cybernética foi ter se empenhado em descrever "o mundo como uma coleção de redes. O segundo é sua intuição de que não existe uma distinção tão clara entre pessoas e máquinas quanto alguns gostariam de crer", sendo o corpo apenas um computador de carne que executa vários *sistemas de informação* que se autoajustam em resposta aos outros sistemas e ao seu ambiente. Nesta visão abstrata de corpo, ele seria similar à internet, pois se constitui por uma coleção de redes. Para tornar esse sistema reticular mais elaborado, otimizando sua funcionalidade, basta reconstruí-lo como um corpo melhor, melhorando seus mecanismos de *feedback* e conectando a ele outros sistemas, como um coração artificial ou um onisciente olho biônico, por exemplo (Haraway; Kunzru;

Tadeu, 2009, p. 126, 125). Vale o destaque: os princípios da maquinização humana já estavam lançados pela Cybernética na década de 40.

Falar de maquinização humana pressupõe entender o conceito epistemológico de máquina, a começar por notar que falar de máquina é *eliminar mistérios*, já que dentro dela não há segredos. A máquina é *projetada* previamente para que se saiba, de antemão, *como* ela vai funcionar, isto é, antes de construí-la, já se sabe *como* ela funcionará sem fazer qualquer experiência prévia disso. E pelo projeto também é possível saber *como* consertá-la se ela estragar. "*Por eso el modelo máquina ejerce una gran fascinación intelectual, ya que si tenemos que habérnosla con un campo de investigación mal conocido podemos proponer un modelo máquina correspondiente y, entonces, todo queda esclarecido*" (Agazzi, 1998, p. 29). E a tecnologia tem ampliado esta forma de lidar com o mundo, pois, a partir dela, elabora-se um *modelo-máquina* para quase todas as coisas.

> *De ahí que, a partir de este momento histórico, la máquina vuelve a ser un modelo teórico reconocido. La mayoría de ellas se pueden transformar en experimentos mentales, o sea, se interpreta una situación compleja según "mecanismos", y hablamos así de los mecanismos psíquicos, de los mecanismos del mercado, etc. La máquina se presenta como un ideal*. (Agazzi, 1988, p. 29)

Importa, no entanto, entender como o emprego do *modelo-máquina* a todos os âmbitos da vida humana pode ser desastroso, já que projetar o modo *como* humanos vão funcionar segundo um padrão significa manipular comportamentos, restringir sua liberdade. Dentre as teses mais reconhecidas sobre a relação máquinas e humanos, encontramos, no pensamento de Deleuze e Guattari, um suporte invulgar a partir do conceito de *servidão*. Segundo eles, dentre as *dinâmicas de existencialização* (Guattari), está a servidão, forçada ou voluntária; e a novidade que aqui nos interessa é a denominada *servidão maquínica* (Deleuze; Guattari, 2011), um novo regime produtivo que *desestabiliza* as representações e comportamentos dos indivíduos, o que os leva a perderem sua *singularidade*. Como tal ocorre? Este regime de produção induz os indivíduos a se comportarem como parte de *servomecanismos* na forma de *agenciamento* (acoplamento de um conjunto de relações materiais e de um regime de signos correspondente a elas) do cotidiano de consumidores oprimidos num sistema virtual de estratégias tecnológicas que conformam a *interação* de humanos e não humanos (Lazzarato, 2014).

Nesse ponto, importa-nos destacar outra referência indiscutível de fronteira, atrelada ao processo de *maquinização/humanização* descrito: a *algoritmização* e a *anonimização* da vida humana, uma implicada na outra. A algoritmização do cotidiano dos humanos é algo recente e sem precedentes na trajetória humana. Na verdade, o cérebro humano é uma montagem de *algoritmos orgânicos* modelada pela seleção natural em milhões de anos de evolução, de modo que a clássica noção de *livre arbítrio*, base de boa parte das escolas de pensamento ocidentais, hoje tem sido desconstruída em prol de teses que mapeiam o cérebro humano e apostam que seu produto (a mente) nada mais é que um complexo e intenso processo de *cálculos* executados, sem percepção da consciência, por milhões de neurônios cerebrais (*algoritmos bioquímicos*) que computam todas as probabilidades que resultam na sobrevivência e na preservação do *homo sapiens* (Harari, 2016, p. 73). A questão é que este mapeamento despertou menos o interesse em investir nele do que o reproduzir como máquina: na década de 1980, embora já contássemos com um maior conhecimento sobre o cérebro, "ele se tornou irrelevante para a nova geração de pesquisadores de IA, cujo objetivo era escrever um programa equivalente, em termos práticos, aos processos cerebrais" (Sejnowsky *apud* Rodrigues, 2021, p. 27).

Falar em algoritmos se tornou o bordão desta era – dita *era pontocom* (*dotcom era*) –, o que, sem compreensão exata do que seja, passou a ser explorado até em *marketing* de perfume. Algoritmo se tornou um termo da moda a indicar que é uma forma melhor, mais atualizada de fazer qualquer coisa, tornando tudo mais inteligente (*smart*), preciso e ágil. Mas sob o ponto de vista estritamente técnico-funcional, algoritmos são apenas *descrições padronizadas de comportamentos*, expressos em termos de um conjunto finito de ações, de modo que, ao se executar a operação "a + b", percebe-se um mesmo padrão de comportamento, ainda que a operação seja realizada para valores diferentes de *a* e *b* (Ziviani, 1999, p. 1). Evidentemente que tais padronizações empregadas no âmbito das relações sociais geram esquemas comportamentais igualmente padronizados, possibilitando o *monitoramento* e a *manipulação* dos indivíduos. A padronização dos humanos em modelos-máquina é exponencial quando realizada por algoritmos.

A mais importante estrutura *relacional* formada por algoritmos é a rede mundial de computadores (*internet*), na qual eles (os algoritmos) estão nos bastidores, monitorando e influenciando o comportamento dos indivíduos, prevendo suas necessidades e ações (Hoffmann-Riem, 2016; Drösser, 2016). Encontramos, na atualidade, vasta literatura sobre *governamentabilidade*

algorítmica e suas perversas amarras à liberdade humana, reduzindo nossa própria condição de humanidade, amarras estas estabilizadas no *big data* e no *data mining*. Elas passam a neutralizar aquelas características tidas por exclusivamente humanas, padronizando subjetividades e criando *standards* comportamentais e relacionais artificiais, de modo a não só excluir perfis fora da média dos padrões estabelecidos, mas também a inadmitir *imprevisibilidades*, o que é próprio das experiências racionais e livres dos humanos. Mais: nos espaços ditos virtuais, comportamentos humanos são moldados segundo padrões que erigem *perfis supraindividuais*, sem deixar espaço para que os indivíduos tenham realmente a noção do que são, com toda a carga de responsabilidade exigida ao *agir ético*, tal como encontramos nos estatutos da Ética tradicional (Vaz *apud* Brochado, 2021, p. 41). Indivíduos passam da condição de singulares a perfis *anônimos* gerais que influenciam outros perfis, pois a padronização é consequência do esquema algorítmico.

Num sistema de cálculos para padronizar formas de vida e de comportamento, é intolerável a reflexão, a recalcitrância, a crítica, elementos que constituem a *subjetividade real* (Berns; Rouvroy, 2013, p. 173-174). Os modelos comportamentais segundo os quais os indivíduos são conduzidos por funções acionadas por algoritmos, que se tornam tanto mais eficazes quanto mais se utilizam dispositivos tecnológicos, substituem a antiga subjetividade, noção constituída por *corpos* e *consciências* individualizados, por perfis automáticos formados a partir de caracteres deixados em milhões de dados diariamente (Berns; Rouvroy, 2013). Aqui vale uma advertência: o sistema mundial de computadores não pode ser tratado como *mais um domínio técnico-científico* sobre o qual se espera que desenvolvamos competências para administrá-lo tal como se encontra, como as agendas da economia global ou do meio ambiente. Nesse ponto, fazemos coro a Morozov em sua advertência quanto ao *pseudoempoderamento do usuário* sobre as redes, ainda que suas posições radicais *antissilicianas* sejam sua vitrine acadêmica:

> Em vez disso, necessitamos de assuntos mais palatáveis – "privacidade" ou "subjetividade", para superar a ideia de rede. Temos de pôr de lado os objetos ambíguos como "liberdade na internet" porque se trata de uma ilusão que não vale a pena perseguir. O que nos compete agora é criar ambientes nos quais a verdadeira liberdade ainda possa ser fomentada e preservada. (Mozorov, 2018, p. 136)

O princípio que rege o *big data* é o acúmulo do máximo de dados, de origens e tipos diversos, a partir da análise dos quais se possam gerar re-

sultados disponíveis para serem amplamente usados em diversos contextos, vale dizer: dados nada mais são do que representações da vida cristalizadas em *quilobytes* (Morozov, 2018, p. 54). E a cristalização da vida humana exige controle, razão por que um dos desafios que se impõe ao Cyberdireito hoje é, além de estabelecer limites à agregação de dados, "criar um equilíbrio entre os interesses dos grandes usuários de dados e aqueles que podem ser adversamente afetados por aplicações de *big data*" (Hoffmann-Riem, 2021, p. 84-85). Conforme registra Hoffmann-Riem, Ministro do Tribunal Constitucional Federal Alemão, as decisões deste Tribunal trouxeram novas formulações sobre o conceito de *direitos fundamentais*, avançando o sistema jurídico alemão, como o direito fundamental à *autodeterminação informacional*, o qual inclui o *direito individual de decidir* sobre o uso e divulgação de dados pessoais, o direito fundamental à garantia da *confidencialidade* e a *integridade* dos sistemas de tecnologia da informação, o que impõe a funcionalidade *técnica e social* a tais sistemas como requisito para a sua utilização autônoma; de modo que "os sistemas de informática protegidos inclu[a]m não só os computadores utilizados pelos próprios interessados, mas também os sistemas de informática que funcionam em rede com computadores externos, por exemplo, quando se utilizam as chamadas *Clouds*" (BVerfGE 120, 274, 313, 141, 220, 264f.; 220 ss., 330 ss. *apud* Hoffmann-Riem, 2021, p. 50).

As pesquisas mais consistentes atualmente nas áreas de humanas sobre o monopólio das experiências humanas por sistemas algoritmizados, tais como as de David Lyon, Shoshana Zuboff e Antoinette Ruvroy, apresentam-se mais como cotejo da *gestão política* de dados disponíveis sobre os indivíduos e a manipulação deles do que propriamente como reflexões sobre o ocaso a nós imposto pela *regulação algorítmica*, que pretende nos vender a ideia de simplificação da vida propiciada pela coleta e o aperfeiçoamento massivos de dados. Facilitaria nossa existência delegarmos um sem-número de tarefas a algoritmos que, "avaliando os resultados de tarefas anteriores e quaisquer alterações nas predileções individuais e nas curvas de indiferença, se reajustariam e revisariam suas regras de funcionamento" (Morozov, 2018, p. 139), costumizando-se às nossas demandas e tornando nossas decisões mais fáceis e satisfatórias. Ocorre que o efeito mais devastador da delegação das deliberações triviais da vida à algoritmização é a anonimização do indivíduo, tanto face a uma totalidade de relações artificiais sem elementos reais de experiência, quanto face à própria *impotencialização* subjetiva perante si memo. E as pesquisas sobre governança algorítmica

mencionadas não aprofundam nesta nova modalidade de *existencialização* dos humanos (Deleuze; Guattari). Elas se ocupam mesmo é das consequências *pragmáticas* do universo *cyber* – o que é legítimo e inovador, mas não exaustivo, visando a detectar os meandros da estruturação desta rede incontrolável de dados e a denunciar suas manipulações imediatas. Todavia, é importante buscar para além destas a mediação que este novel *processo relacional* dos humanos provoca em seus códigos comportamentais e existenciais, já que as crenças dela decorrentes impactam nossa *existencialização* presente, permeada por *códigos virtuais maquínicos* que implicam extrema maquinização da experiência humana, jamais vista até aqui.

Certamente o enfoque destas pesquisas se justifica pela necessidade urgente de mapear os processos e tentar dominar seus efeitos, pois elas enfrentam a difícil tarefa de criar um *modelo-máquina de governança* para controlar esse outro gigantesco modelo-máquina que é o advento *big data*, e sobre o qual não se tem o controle. E não o tem porque o projeto inicial não traz soluções nem para as suas disfuncionalidades nem para o seu produto, hoje inadministrável e desconhecido, já que nós humanos produzimos uma quantidade de dados que hoje ultrapassa 1 *zettabyte* (equivalente a 1 sextilhão de *bytes*) e não sabemos o que será efetivamente feito disso. Sem mencionar que a anomização dos sujeitos habitantes desse universo constituído de dados toca numa questão de especial interesse para o Direito: a privacidade, que vem se tornando uma mercadoria. Como também adverte Morozov,

> a privacidade deixou de ser uma garantia ou uma coisa de que desfrutamos gratuitamente; agora temos que gastar recursos para dominar as ferramentas. Esses recursos podem ser dinheiro, paciência, atenção – dá até para contratar um consultor que se encarregue de fazer tudo isso –, mas a questão é que a privacidade hoje é algo caro. (Morozov, 2020, p. 36)

A questão ética central que se impõe a nossa reflexão não é de ordem técnica ou funcional sobre um mundo maquinizado e datificado. O dilema ético que testemunhamos impotentes é não conseguirmos nos aperceber como fantoches num processo de manipulação que considera nossos desejos e aflições perante a vida nada mais que dados rentáveis a médio prazo – hoje a base do *capitalismo dadocêntrico*. Não menos relevante é estarmos atentos ao fato de que, ao mesmo tempo em que somos datificados, as práticas *facilitativas* a nós disponibilizadas na rede nos transformam em seres dependentes e carentes de "dicas" para praticar as ações mais banais da vida,

o que não era possível há poucos anos, quando lucidamente acreditávamos que a dúvida fazia parte das descobertas próprias da nossa existência.

O sistema pseudodemocrático da algoritmização (orientada por inteligência artificial) incutiu em nós a crença em *explicações monocausais* das ocorrências no mundo e esta crença tem nos involuído enquanto capacidades pensantes criativas, instalando uma apatia generalizada nas gerações mais jovens. Estas parecem não mais se interessar por questões complexas e muito menos valorizar problematizações e eurecas – experiência denominada *thauma* (θάμα) pelos gregos, que significa a capacidade humana de admiração e espanto diante de uma descoberta, o que nos traz a conquista da *eudaimonia* (autorrealização na razão). Aqui lembramos o singelo e notável aviso de Martin Hilbert, intelectual alemão pioneiro em pesquisas que quantificaram o número de informações existentes no mundo hoje: "Nos preocupa muito ver nossos filhos grudados o dia inteiro em uma *chupeta digital*, incapazes de se concentrar ou assimilando expectativas pouco realistas sobre seus corpos. Mas nós somos outra coisa, usamos as redes por diversão, ninguém está colocando uma chupeta na nossa boca" (Hilbert, 2018). Essa "chupeta virtual", na ilustração de Hilbert, poderia ser apontada como o incômodo refletido nas colocações trazidas neste trabalho e que nos indica uma segunda conclusão parcial: quatro são as premissas investigativas que orientam as próximas questões, quais sejam, as peculiaridades da *cybervida*, o *novo* humano, a *preocupação* da Filosofia e a *missão* do *Direito* nessa processualidade presente.

6 SOBRE HUMANIZAÇÃO DA MÁQUINA OU INTELIGÊNCIA ARTIFICIAL FORTE

É segundo as premissas lançadas na conclusão parcial trazida no tópico anterior que acreditamos ser essencial dar um passo para além das pesquisas sobre *governança de dados* e tentar chegar a mais alguns enclaves sobre outro advento marcante que está ligado à maquinização e à algoritmização: a *humanização da máquina*. A independentização das faculdades humanas (de conhecer e raciocinar) das condicionantes impostas pelo meio levou os humanos à apoteótica experimentação de si, projetando-se em *máquinas inteligentes*, construindo seres dotados de *inteligência artificial* (IA), em alguns casos literalmente a sua imagem e semelhança. Um deles ficou badalado na mídia e tem perfil com milhares de seguidores no *Instagram*: a *humanoide* Sofia, criada pelo roboticista David Hanson, uma arquitetação exemplar de *humanização de máquina*. Inspirada na atriz Audrey Hepburn, Sofia foi projetada com aparência humana refinada, capacidade de dialogar

e formular questões complexas, manifestar 62 expressões faciais, dentre outras manifestações exclusivamente humanas. Dentre seus feitos está a maior aproximação jurídica com os humanos: o reconhecimento da sua condição de cidadã saudita.

Na verdade, a espécie *homo sapiens* evoluiu até realizar a proeza de reproduzir, com maestria, a base (e sentido) de sua própria existência, o cérebro humano, em formatos artificiais diversos, desenvolvendo o campo da chamada *computação cognitiva*. As tarefas desta são engendradas por milhares de algoritmos estruturados como fórmulas cada vez mais refinadas que possibilitam a *aprendizagem de máquina* (AM), que, pelo grau de refinamento *preditivo*, são classificadas como *machine learning* e *deep learning*. Estas inteligências artificiais adquirem conhecimento extraindo padrões a partir de dados *não trabalhados* (Ziviani, 2017), isto é, são programas de computador capazes de aprender a executar tarefas a partir de sua própria experiência (Faceli *et al.*, 2011). Trata-se de sistemas de aprendizagem algorítmica capazes de se adaptar a novas situações problemáticas de forma independente, seguindo escrevendo seus próprios programas (Hoffmann-Riem, 2021, p. 15). Vale dizer: *algoritmos de aprendizagem* não são programados para resolver problemas específicos, mas, sim, para aprender a resolver problemas, fazendo *predições*, por exemplo, sobre qual filme ou marca de sabonete uma pessoa pode gostar, de modo que algoritmos aprendizes são aqueles que fazem outros algoritmos (Tutt, 2017, p. 85).

A predição é um processo por meio do qual a máquina se socorre de dados, que nada mais são que informações, para preencher lacunas informacionais, criando novas informações destas combinações de informações, isto é, a dita *capacidade de predição da máquina* é a organização que ela faz

> autonomamente [d]os dados de entrada para predizer [a] melhor solução de saída (a exemplo da navegação pela ferramenta Waze, na predição do melhor caminho a ser tomado, frente o *data driven* de um contexto específico), em superação a um modelo em que a solução estava vinculada a uma programação previamente codificada (a exemplo do que se verificava na lógica do GPS, preso a soluções indicadas por um mapa previamente codificado). (Rodrigues, 2021, p. 24)

Esta nova forma de projeção da mente humana é o maior desafio que se coloca nessa quadra de século sobre os limites do *humano* e do não *humano*, já que temos máquinas muito mais inteligentes, eficientes e velozes

que humanos, e, portanto, potenciais substitutos deles em vários setores da sociedade. Essa afirmação é perigosa para a Filosofia e para o Direito, áreas que se equilibram sob o suposto da razão e da liberdade entendidas desde sempre como adjetivos *exclusivamente* humanos. Como pontua Yuval Noah Harari, se o problema do século passado foi a *exploração* de humanos por humanos, o deste século é a absoluta *irrelevância* de humanos (Harari, 2016), ou seja, o que em nós até aqui era supervalorizado, nossa capacidade de associação, criação e memorização, está fadado ao descaso pelo capitalismo dadocêntrico, esta nova feição de sistema econômico que transforma todos os dados da existência humana cotidiana em ativo rentável (Morozov, 2020, p. 33).

Teses filosóficas clássicas, como as teorias biológicas do conhecimento (Macht e Avenarius) e as que defendem a *lei da heterogenia dos fins* (Vaihinger), já se ocupavam da hipótese de redução da racionalidade humana a processos biológicos e filogenéticos; a novidade do final do século passado, intensificada de forma galopante neste, são os avanços das pesquisas em neurociências e ciências da computação, as quais permitiram a laminação e mapeamento do cérebro humano, bem como a reprodução precisa de suas atividades e produções. Nivio Ziviani aponta o curioso exemplo da criação de músicas por inteligência artificial que imitam perfeitamente o *estilo* de um compositor falecido, o *rapper* Sabotage.

A questão mais essencial que se coloca a uma Filosofia Algorítmica é: máquinas de altíssima *performance* cognitiva não são dotadas de liberdade tal como concebemos por séculos, mas realizam tarefas a partir de aprendizado célere e infinito, o que pode ser caracterizado como racionalidade independentemente de estar atrelada à consciência. E, se dissolvemos esse liame, o conceito de liberdade fica prejudicado, eis que nos acostumamos a associar a faculdade de *conhecer* à faculdade de *agir* livremente, aspectos imbricados da experiência humana. Conhecer algo sem ter consciência de que se está conhecendo soa *non sense*; e, tentando superar propostas antediluvianas já estabilizadas na área jurídica (especialmente a noção de imputação como decorrente da experiência consciente), precisamos dar um passo à frente ao nos propormos a refletir sobre o impacto da dissociação entre *inteligência* e *consciência* para o Direito e suas engrenagens coercitivo-punitivas. É dizer: olvidar esta dissociação no estágio atual das experiências humanas é inércia intelectual insustentável para os juristas, pois é o suposto de toda a experiência jurídica e temos que nos ver com o fato de que máquinas aprendizes tomam decisão, em que pese estes sistemas de

dados não conseguirem desenvolver qualquer narrativa sobre a realidade sob pontos de vista histórico ou ideológico.

E é exatamente neste aspecto que os criadores dos vários sistemas de IA forte maquiam a deficiência da monocausalidade explicativa do real e vendem-na como *objetividade* (Morozov, 2018, p. 141), característica decisional avidamente desejada no âmbito da atividade jurisdicional, não só pela carga de segurança jurídica que fingem oferecer, mas também pela desincumbência quanto à carga moral da decisão (Habermas) por sujeitos falíveis, os juízes humanos dotados de consciência – algo que "pesa". De modo que a clássica noção de julgamento *equidoso* tem sido algoritmizada e anonimizada, mais um produto do modelo-máquina do *big data*, e que esconde dos juízes o principal desfecho: eles também, em breve, terão suas subjetividades absorvidas por padrões que dissolvem as singularidades em perfis automáticos formados por milhares de dados, tornando-os igualmente servomecanismos (Deleuze; Guattari). Argumentam os membros da *ITechLaw* e seus seguidores que a IA está sendo e será usada para atividades repetitivas, mas esta é uma definição difícil quando tratamos de questões humanas, pois nada se repete exatamente quando falamos de experiências singulares, e obviamente que, se uma IA forte puder ser manejada para buscar melhores soluções, não haveria razão para não a empregar: imaginemos, no âmbito do direito de família, como soaria aos interessados saberem que suas demandas seriam julgadas por máquinas inteligentes hiperpotentes. Isso estafaria as bases institucionais do Judiciário em pouco tempo, sem qualquer tom escatológico, pelo malogro nos escopos do próprio *Poder* Judiciário, o qual cumpre a missão de receber demandas e decidi-las com competência e sensibilidade (não à toa sentença vem de *sentire*), no que reside simbolicamente sua legitimidade, respeitabilidade, sua posição de *poder*. Pessoas seguirão buscando por autoridades para resolver com sensibilidade humana seus dramas pessoais quando sabedoras serão de que robôs estão à frente dos processos? Parece que será algo similar ao que os *chatbots* provocam em nós hoje: busca por empatia ao tentarmos chegar num humano que nos atenda e resolva nossas questões *humanamente*.

7 SOBRE A COMPOSIÇÃO MÁQUINA-HOMEM E SUA RADICALIZAÇÃO: O TRANSUMANISMO

Para além da abundante exploração das discussões técnicas do universo *tech*, importa conceder que a interpretação filosófica é que nos permite ousar propor novo enfoque sobre seus platôs algoritmizados, o que aqui chamamos *filosofia algorítmica*. Este enfoque pretende ser, como toda

proposta filosófica, uma *reflexão de terceiro grau* (como definido no início deste texto) e, para tanto, precisamos partir de teses já consagradas na história da filosofia ocidental, tal como a tese sobre *governamentabilidade* de Michel Foucault, já amplamente acessada nas discussões sobre governança algorítmica, e a leitura *rizomática* de Deleuze e sua compreensão de agenciamento e acoplagem humano-maquínica, e que chega a tratar a questão com certa indulgência. Segundo o filósofo,

> já não se trata de confrontar o homem e a máquina para avaliar as correspondências, os prolongamentos, as substituições possíveis ou impossíveis entre ambos, mas [trata-se de] levá-los a comunicar entre si para mostrar como o homem *compõe peça com* a máquina, ou compõe peça com outra coisa para construir uma máquina. (Deleuze; Guattari, 2011, p. 508)

Na mesma esteira crítica, não podemos deixar de registrar as audazes críticas à epistemologia tradicional feitas por Bruno Latour ao anunciar a *hibridização* ofuscada pela longa trajetória da filosofia legatária da *modernidade* que não dá conta de responder a uma constatação: os híbridos existem e são fruto de formas de conhecimento não catalogáveis pelos modelos binarizantes das ciências. Todos os dualismos no qual acreditamos por séculos "foram canibalizados ou, como diria Zoe Sofia (Sofoulis), eles foram 'tecnodigeridos'. As dicotomias entre mente e corpo, animal e humano, organismo e máquina, público e privado, natureza e cultura, homens e mulheres, primitivo e civilizado estão, todas, ideologicamente em questão" (Haraway; Kunzru; Tadeu, 2009, p. 63). Latour é veemente ao criticar a hermenêutica "inventada" pelas ciências humanas e que deixou "o mundo das coisas derivar lentamente em seu vazio" ao mesmo tempo em que os cientistas e tecnocratas expandem seu território laboratorial naturalista. Tudo o que a empreitada epistemológica conseguiu promover desde o advento da modernidade foi radicalizar cegamente cisões epistemológicas que não mais se sustentam face aos *híbridos* que surgiram ao largo e em afronta a elas, de modo que se tornaram caricatas divisões do mundo, tão caricatas que, num mesmo projeto civilizacional de descoberta de verdades científicas, encontramos neurocientistas descrevendo neurônios de um lado e psicanalistas fazendo análise de estados psicopatológicos de outro como se não estivessem tratando de uma mesma realidade, sem dialogarem entre si e sem se darem conta de seus recortes artificiais falaciosos (Latour, 2013, p. 59).

A hibridização não é apenas de ideias, pois ideais encarnam-se em coisas; a hibridização passou a ser um projeto *ideológico* dos atuais hu-

manos. E, ao afirmar a potencialidade da filosofia em se tornar *reflexão algorítmica,* atribuímos a esta forma de pensar a busca de resposta para uma macroquestão de fundo: com algoritmos no *comando* (lembremos que cybernética é comando) do funcionamento das relações/inteirações humanas independentes de *corporeidade*, quais direitos e deveres fundamentais podem advir dessa nova experiência que independe de estrutura biopsíquica? A pergunta é dirigida à missão jurídica que lançamos como uma das premissas desse trabalho, mas as razões da indagação são de natureza filosófica, e a radicalização da vivência humano-maquínica é apresentada à filosofia algorítmica como *movimento e doutrina trans-humanista*. Trazer algumas referências dela é cogitar de um modelo-máquina para enfrentar outras questões que se aproximam mais ou menos dela, já que ela é um modelo radical de lidar com a situação humana atualmente e com vistas a um futuro que possa realizar o projeto transumanista.

O movimento *trans-humanista* tem seu próprio manifesto, com valores, direitos e deveres declarados, e sua face mais questionável eticamente é a compreensão, por seus seguidores mais radicais, de que a corporeidade humana constituída de carne (corpo biológico) é abjeta e deve ser substituída (ou ao menos exponencialmente melhorada) para se libertar dos condicionamentos naturais (Bostrom, 2005; More, 1990; Young, 2005). A palavra trans-humanismo (ou transumanismo) foi adotada pela primeira vez em 1957 por Julian Huxley (sim, irmão de Aldous Huxley), que, além de reconhecido cientista da área de Biologia, foi o primeiro diretor-geral da Unesco e fundador do *World Wildlife Fund*. Na obra *New bottles for new wine*, ele escreveu que deveríamos buscar um meio de superar nossas limitações físicas e nos transcendermos a partir de nossa natureza humana, definida pela nossa mente singular. E assim o termo surgiu:

> *The human species can, if it wishes, transcend itself – not just sporadically, an individual here in one way, an individual there in another way – but in its entirety, as humanity. We need a name for this new belief. Perhaps transhumanism will serve: man remaining man, but transcending himself, by realizing new possibilities of and for his human nature*. (Bostrom, 2005, p. 7)

Desde então, o transumanismo tornou-se um movimento que defende o emprego da biotecnologia para promover a adaptação de seres humanos e reduzir os riscos existenciais, melhorar a saúde, a memória, a longevidade, o bem-estar e a prosperidade humana (Bostrom, 2005, p. 25). Com várias versões, inclusive com vieses ficcionais, o movimento se consolidou como

projeto acadêmico com a fundação da *World Transhumanist Association* em 1998 por Nick Bostrom e David Pearce. Este publicou *The Hedonistic Imperative*, em que defende a intervenção neurotecnológica para eliminar o sofrimento dos "animais humanos e não humanos". A associação transumanista publicou a Declaração Transumanista e promove debates acalorados na internet, nos quais transumanistas e *bioconservadores* discutem suas convicções, bases teóricas e propósitos, além de organizarem um evento anual intitulado *TransVision* e publicarem o jornal *on line Journal of Evolution and Technology* (Bostrom, 2005, p. 16). Chama a atenção entre os princípios da Declaração Transumanista atualizada em 2009 o seguinte:

– Defendemos o bem-estar de todos os sencientes, incluindo humanos, animais não humanos e quaisquer futuros intelectos artificiais, formas de vida modificadas ou outras inteligências às quais o avanço tecnológico e científico possa dar origem.

– Somos a favor de permitir aos indivíduos uma ampla escolha pessoal sobre como eles capacitam suas vidas. Isso inclui o uso de técnicas que podem ser desenvolvidas para auxiliar a memória, a concentração e a energia mental; terapias de extensão de vida; tecnologias de escolha reprodutiva; procedimentos criônicos; e muitas outras modificações humanas possíveis e tecnologias de aprimoramento.

A declaração não só defende modificações humanas com o emprego de tecnologia de aprimoramento, mas também defende o bem-estar de intelectos artificiais e quaisquer formas de vida modificadas que o avanço tecnológico possa desenvolver. Pode parecer ficcional demais para ser levado a sério num artigo científico, mas esse tipo de manifestação nos alerta para a urgência de discutirmos a maquinização humana para além dos seus feitos e resultados, o que não pode se restringir ao âmbito do emprego de mecanismos sofisticados que facilitam a vida e trazem mais conforto para os humanos. As novas gerações de humanos estabelecem uma relação tão visceral com máquinas (em razão do refinamento funcional das mesmas), que passam a considerar as limitações do *corpo biológico* intoleráveis.

Curiosamente a Filosofia ocidental racionalista sempre pregou que o diferencial do humano era a sua racionalidade (homem = animal racional) e sempre considerou o corpo um fardo carregado, suportado pelo espírito. Ocorre que a concepção de *elevação do espírito* era uma exortação à transcendência da alma/intelecto sobre as limitações físicas, buscando o

desenvolvimento das potencialidades intelectuais e morais, que não deveriam sucumbir às afecções do corpo, dos desejos e instintos. Mas o que está por trás do movimento trans-humanista não é singelo assim, na forma que os filósofos idealizavam a elevação do espírito humano sobre a decrepitude do corpo. É algo muito diverso e radical: é a convicção de que é preciso *se livrar* do próprio corpo (biológico) e encontrar formas *híbridas* de existência, de modo a transferir o máximo de funções corporais para máquinas.

A perspectiva trans-humanista concebe o corpo como algo desprezível, submetido a toda forma de condicionamento natural, desde doenças, sofrimentos psíquicos decorrentes de ausência de determinadas substâncias cerebrais, até o irreversível envelhecimento que nos leva irremediavelmente à morte Encontramos afirmações curiosas entre pesquisadores transumanistas, como o filósofo Antonio Diéguez, que chega a lamentar que "o problema do ser humano é estar num *suporte* errado" (Disponível em: http://www.ihu.unisinos.br/78-noticias/594942); é dizer: o projeto trans-humanista crê ser possível buscar um suporte mais adequado para a mente, a qual poderia ser transferida para um disco rígido (o chamado *mind uploading*). E essa convicção encontra respaldo em pesquisas e instituições, como o caso da famosa *Alcor*, empresa de *criogenia* instalada no Arizona (EUA) que cobra algo em torno de oitocentos mil reais para congelar um corpo humano e cem mil para congelar uma cabeça.

Não estamos aqui escrevendo comentários a romances ficcionais no estilo de Arthur C. Clarke ou Aldous Huxley; estamos relatando manifestações e convicções de indivíduos reais do nosso presente e que creem na tecnologia para a realização desses fins (inclusive pagando por ela). E o grande problema é que estas convicções são, de certa forma, o que se anunciou como projeto de toda a Filosofia: a insubmissão da mente às vicissitudes causais impostas pelo corpo. Nick Bostrom sugere que a grande inspiração para o transumanismo foi a filosofia de Nietzsche e seu conceito de super-humano (*Übermensch*) trazido na obra *Assim falou Zaratrusta*, mas a tese nietzschiana, como alerta o próprio Bostrom, referia-se ao crescimento pessoal e refinamento cultural de certos indivíduos, especialmente os que conseguiam superar a moralidade escrava do Cristianismo. O que Nietzsche tinha em mente, entretanto, não era uma transformação tecnológica, mas "uma espécie de crescimento pessoal e refinamento cultural elevados em indivíduos excepcionais" (Bostrom, 2005, p. 6). A apropriação desse antigo tema da Filosofia por arenas tecnocientíficas deve ser preocupação da Filosofia Algorítmica.

8 SOBRE FILOSOFIA ALGORÍTMICA E CYBERDIREITO[4]

Ao apontamos o Direito como disciplina *Cyber*, compreendemos essa afirmação duplamente: no sentido de sistema regulador que é, e no sentido de ter de atender ao chamado de regular adventos ancorados num mundo em que os desdobramentos tecnológicos da Cybernética impactaram as relações humanas que se configuram como direitos e deveres. É possível detectar esta percepção do problema em pontos de vista de juristas contemporâneos, embora eles empreguem genericamente a palavra "tecnologia" para indicar a função e os novos desafios postos para o Direito.

Tércio Ferraz Junior é um dos que defende as funções tecnológicas (eu diria cybernéticas) do Direito. Conforme explica, o Direito enquanto ciência dogmática cumpre "funções típicas de uma *tecnologia*. Sendo um pensamento conceitual, vinculado ao direito posto, a dogmática pode instrumentalizar-se a serviço da ação sobre a sociedade" (Ferraz Júnior, 2003, p. 85). Nos moldes do *modelo maquínico*, o Direito cria um esquema de solução pautado por entradas, saídas e calibrações, de modo que a dogmática é acima de tudo *controle*,

> na medida em que seus corpos doutrinários delimitam um campo de solução de problemas considerados relevantes e cortam outros, dos quais ela desvia a atenção. [...] Nesses termos, um pensamento tecnológico é, sobretudo, um pensamento fechado à problematização de seus pressupostos – suas premissas e conceitos básicos têm de ser tomados de modo não problemático – a fim de cumprir sua função: criar condições para a ação. No caso da ciência dogmática, criar condições para a decidibilidade de conflitos juridicamente definidos. (Ferraz Júnior, 2003, p. 85)

Não por outra razão é que, dentre os recortes das disciplinas críticas do Direito, temos numa *teoria jurídica* do *controle de comportamentos*, a qual cuida da organização jurídica do exercício do poder e dos mecanismos

4 Não adotamos o neologismo cyberdireito no sentido de *juscibernética* proposto por Mario Losano, que, evoluiu o conceito de *jurimetria* pensado desde a década de 40 por Lee Loevinger. Este tentou conformar uma nova ciência empírica (que ele nomeou jurimetria) voltada para de três âmbitos: a elaboração eletrônica de dados jurídicos; a aplicação da lógica computacional na área do Direito; e a análise comportamental das práticas nos Tribunais. Losano teria pretendido ampliar a jurimetria com uma juscibernética, voltada ao "estudo do sistema jurídico, da sua estrutura dinâmica, dos elementos que o compõe e das técnicas, mediante as quais, um problema de Direito pode vir a ser tratado pelo computador eletrônico" (CARRAZZA, 1974, p. 60). Estas teses buscaram trazer os feitos cibernéticos, basicamente a evolução das técnicas computacionais, para a solução de problemas na área jurídica. Adotamos cyberdireito não nesse sentido, mas para designar uma leitura teórica do Direito num mundo *cyberizado* que impacta radicalmente a cultura, as relações sociais, e, evidentemente, o direito, segundo o que nesse artigo se propõe.

políticos que dão efetividade a este exercício enquanto capacidade de provocar obediência nos destinatários do *controle* exercido pelo Direito. Como equacionado por Ferraz Júnior, essa teoria liberta a Ciência do Direito da prática limitada da exegese do sistema jurídico, "como se o Direito fosse apenas um *dado* que competiria ao jurista examinar. Ela vai mais adiante e exige uma concepção do Direito como uma verdadeira técnica de invenção, algo que não está pronto, mas está sendo constantemente construído nas interações sociais" (Ferraz Júnior, 1980, p. 101).

Dispondo deste exemplo de concepção doutrinária, fica fácil notar que há teses no campo da Ciência do Direito que tangenciam a questão da influência da Cybernética sobre o Direito, ainda que não o façam explicitamente. A tentativa de buscar um lugar para esta ciência como estatuto científico privilegiado nos desdobramentos epistemológicos deste século é um desafio que deixamos em aberto. Com efeito, vale indicar que, de certa forma, considerar o Direito um *sistema tecnológico de controle* é admitir sua natureza cybernética, ou seja, ele é uma forma comunicacional de controle e que se enquadra nas definições de regulação e autorregulação sobre as quais se sustenta o diferencial da ciência cybernética como *ciência do controle*.

O ponto de convergência entre filosofia algorítmica e cyberdireito é o advento da dissolução dos sujeitos em padrões consumidos pela adesão *implícita* dos indivíduos a toda forma de manipulação maquínica conduzida por algoritmos, a exigir das ciências humanas a reflexão sobre a situação dos sujeitos num *mundo da vida* (Husserl) *cyberizado* e *algoritmizado*. Tal convicção nos permite afirmar que uma filosofia para o século XXI deve levar em consideração a experiência jamais antes vivenciada pelo *homo sapiens* de ter sua replicação intelectual em cérebros artificiais hiperpotentes e sua individualidade desconsiderada por um sistema de dados que massifica e torna irrelevante a posição dos sujeitos no mundo. Falar de uma Filosofia Algorítmica é assumir que a hermenêutica do tempo presente, aí incluída a hermenêutica jurídica (Betti, 1990), projeta-se sobre um mundo jamais experimentado até o final do século passado. A própria noção de tempo (cada vez mais acelerado) passa a ser representada de outra forma, numa *ductibilidade comportamental* denunciada por filósofos como Zygmunt Bauman e Byung-Chul Han.

Aqui situamos nossa terceira conclusão parcial (haja vista que esse artigo é um esforço de conclusões parciais): as tentativas de conexão entre *fundamentos* da realidade pensados em termos de algoritmos que subsidiam

todas as relações virtuais e *manifestações cyber* desta realidade é a base para a posição *pragmática* que o Direito deve assumir no sentido de avançar para um sistema de proteção de *cyberdireitos* que ultrapasse o atual estágio de *proteção de dados* num esquema de ressarcimento ainda patrimonializado por demais. Tal já se anuncia em documentos internacionais, como a *Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente* (CEPEJ/31ª Reunião Plenária-Estrasburgo/2018) e as cláusulas 36 e 37 da *Declaração de Bávaro*, além de outros avanços com enfoque menos financeirista e mais ético-preventivo, como o *Comunications Decency Act of 1996* (EUA). Releva anotar também o surgimento de novas áreas da Ciência do Direito, que têm sido nomeadas Direito Digital, Direito da Inovação, Direito Robótico, Direito Orientado a Dados etc. Essas áreas compõem o mosaico de discussões que têm um enredo de fundo comum: a *regulação* de novos direitos decorrentes da situação dos humanos, também em perspectiva hibridizada, enquanto sujeitos de direitos fundamentais num mundo *cyber*, onde o controle é exercido por *algoritmos*.

Nesse contexto de controle *meta-humano* dos próprios humanos, a Ciência do Direito se vê diante de um paralogismo indiscutível: a ausência de liberdade impede a *culpabilização* e a responsabilização de entidades que não gozam da condição tradicionalmente considerada humana. Não se trata aqui de amesquinhar a discussão em modalidades sancionatórias do tipo: é possível impor indenização a atividades realizadas por uma parafernália tecnológica qualquer, tendo ou não caracterização humanoide, haja vista que ela pertence a alguma pessoa física ou jurídica. Esta é a situação do problema ainda em termos romanísticos. O que pretendemos apontar como discussão que se impõe à Filosofia do Direito e à Hermenêutica Jurídica atuais, inspiradas pelas concepções neste artigo lançadas sobre uma Filosofia Algorítmica a inspirar a nova configuração da experiência jurídica como Cyberdireito, é o fato de que a sede decisional no mundo contemporâneo não se encontra mais enclausurada na exclusividade da consciência humana, base de toda a Filosofia da consciência legada definitivamente por Kant. O conceito de sujeito que se enraizou na Filosofia ocidental desde o *cogito* cartesiano subjaz a todas as teorias sociais e políticas ocidentais, é ele o fundamento da ideia moderna e liberal de democracia, sendo até mesmo a base da ideia moderna de educação. No entanto, se existe

> uma criatura *tecno-humana* que simula o humano, que em tudo parece humana, que age como um humano, que se comporta como um humano, mas cujas ações e comportamentos não podem ser retroagidos a nenhuma *inte-*

> *rioridade*, a nenhuma racionalidade, a nenhuma essencialidade, em suma, a nenhuma das qualidades que utilizamos para caracterizar o humano, porque feita de fluxos e circuitos, de fios e de silício, e não do macio e fofo tecido de que somos ainda feitos, então é a própria *singularidade* e *exclusividade* do humano que se dissolve. (Haraway; Kunzru; Tadeu, 2009, p. 13)

Questionar a singularidade humana em sua interioridade consciente é uma *disrupção* que toca em cheio toda a história do Direito ocidental, não só porque coloca em xeque a condição de *sujeito de direito* enquanto destinatário das normas jurídicas, mas também a diferenciação das *autoridades* que decidem sobre a aplicação das mesmas. O papel desempenhado pelo juiz enquanto *mediador* entre a objetividade da *lex* e a máxima que deve dela decorrer para cada caso submetido à mesma *lex* é o momento de concretização última do fenômeno jurídico (*universal-concreto*). A tarefa do julgador é efetivar a *aequitas*, adequando abstrações à realidade das coisas, dos detalhes de cada caso humano, vale dizer, promover a adaptação (*epiéikeia*) (Brochado, 2021), segundo a qual ele realiza a *eunomía* da norma, já dotada de *isonomía* no plano abstrato (Vaz, 1993, p. 48-49). Sem querer requentar o óbvio dito à exaustão na área jurídica, a *medida* aplicada ao caso necessita de uma consciência humana, e tal medida deve ser proporcional às peculiaridades de cada caso, realizando justiça *diferenciada* para cada *singularidade* face à mesma *fórmula abstrata* e *geral* legal. Não há de se falar em fruição de direitos ou Estado de direito sem o papel exercido pelo Judiciário de *garantidor* da aplicação dos direitos fundamentais, por meio do instrumento da *actio*, com força de *coisa julgada* (Salgado, 2006).

Esta é a cartilha que rezamos desde as revoluções burguesas e ela é, sem dúvida, um legado *humanista*. Num mundo em que se discute trans-humanismo, em que o acoplamento homem-máquina é pano de fundo para muitas discussões e realizações, em que algoritmos substituem humanos, a consciência humana perde protagonismo. Este tipo de constatação levada para o âmbito do funcionamento do Poder Judiciário gera um sério problema de fronteira para o Direito, fenômeno ético que tem seu sustentáculo final na manifestação de *consciências* de autoridades humanas investidas de poder decisional sobre casos submetidos a categorias *objetivas* estabelecidas em lei, o que nomeamos *consciência jurídica* (Brochado, 2021, p. 108-159).

O emprego da inteligência artificial na condução dos processos não é a única questão que gera preocupação. O julgamento por humanos tam-

bém é passível de manipulação algorítmica. As decisões e os erros decisionais podem decorrer não só de julgamentos não dotados de consciência (IA fraca ou forte), como vem sendo amplamente empregado pela atuação da *justiça preditiva* (conformada a métodos bayesianos), mas também de uma forma paralela de *manipulação* dos julgadores humanos por algoritmos: a manipulação do próprio juiz enquanto subjetividade imersa num mundo algoritmizado. Nesse sentido, a preocupação de Bruno Alves Rodrigues sobre a real possibilidade de violação da privacidade de Magistrados

> na busca por padrões de comportamento denotados em redes sociais particulares, de forma a permitir construções cirúrgicas de argumentos, analogias e metáforas eficientes à sensibilização do julgador, como parte de determinada estratégia judicial, [podendo] até mesmo sugestionar as opiniões de um Magistrado, por meio do uso do *marketing* digital direcionado ao patrocínio de determinadas publicações na *timeline* de suas redes sociais. (Rodrigues, 2021, p. 246)

Esta é uma dentre tantas questões que se colocam para o Direito, desde o controle de projetos de robótica na arquitetação de humanizações maquínicas nas modalidades *robô, ciborgue, androide (ginoide), humanoide, humanos digitais* etc., e que se aproximam da *singularidade*, até o enfrentamento do limite a ser imposto entre cidadania robótica (e direitos dela decorrentes) e intolerância humana a máquinas altamente *antropoformizadas* no limiar do *vale da estranheza*. Isso sem mencionar as demandas já banalizadas sobre a circulação descontrolada de *criptomoedas*, o alcance da proteção de terceiros contra *smart contracts*, a responsabilização civil de máquinas, a obsolescência do trabalho humano, visto que o principal impacto da internet 4.0 é a substituição das atividades humanas, além de questões de alta envergadura ética ao se cogitar uma "humanidade 2.0", tal como o sugerido por Eric Dietrich, que vê a superação histórica das limitações éticas dos humanos por máquinas libertas de mecanismos de sobrevivência que ainda se impõem ao cérebro humano de dos quais cérebros artificiais estariam imunes (Dietrich, 2004), já que podem ser projetados segundo uma tabela complexa de Ética das virtudes (Teleológica). No campo da *Ética Computacional*, tem havido o resgate do aristotelismo, de modo a possibilitar surgir uma nova categoria de *agentes morais artificiais* que, inclusive, poderão ensinar e ser modelo comportamental para os humanos, enquanto *seres virtuosos infalíveis* que estas máquinas serão.

O *ativismo tecnológico* é a esteira por onde rola um sem-número de arranjos institucionais, políticos e empresariais, com consequências que im-

pactam o Direito e sobre os quais precisamos urgentemente nos debruçar, particularmente para formar cabedal teórico para subsidiar julgados nos tribunais, o que, no caso do Brasil, é agravado por posicionamentos judiciais convenientemente direcionados ao monopólio *tech*. Um exemplo simples, que tem perturbado os juslaboralistas, é a *plataformização laboral* e a forma como o nosso Judiciário tem lidado com ela. Por escapismos dogmáticos questionáveis quanto à desatualização da legislação brasileira, empresas estrangeiras têm sido exculpadas na imensa maioria dos julgados, apesar de ser visível a exploração de mão de obra dos brasileiros sem qualquer ônus quanto a direitos trabalhistas e previdenciários. Igualmente o Estado tem assumido prejuízos decorrentes desta falta de responsabilização jurídica: citamos o caso de um moto-entregador que teve a perna lacerada em acidente de trânsito no percurso da entrega de comida e o custo do tratamento, mais de quinze mil reais, foi assumido pelo SUS. Sem mencionar que somos lesados no mercado de divisas, já que o lucro dessas empresas de aplicativos nunca é investido em nosso país. O caso do *app-trabalho* é apenas a ponta de um arcabouço de problemas ainda estranho ao Direito e que requer atenção quanto às relações estabelecidas por indivíduos num contexto *fluido* e *intangível*, o universo dito *digital*, não passível de fiscalização e controle concreto de realidades tangíveis tal como se desenhou no esquema coercitivo legado pelo Direito Romano. A pandemia mundial vivida neste último ano só fez radicalizar e acelerar o *ativismo tecnológico*, no que os *techies* deram verdadeiro xeque-mate nos tecnofóbicos e luditas.

Para responder a tantas questões aqui suscitadas, nossa posição final é de que a próxima tarefa que a nós se impõe é a reflexão sobre a mudança de rumos do padrão científico do tempo presente, levando-nos a indagar qual o critério adotado até então para definir o humano e o *in*-humano, o *pós-humano* e o *trans-humano*. Por se situar esta composição de reflexões entre as fronteiras da Filosofia, do Direito e da Cybernética, o desafio de verticalização das ideias aqui preconizadas é apontar a posição da *filosofia do direito* sobre formas relacionais que a dogmática jurídica tem o compromisso social de regular, no que tem sido impotente diante de demandas surgidas como novos direitos fundamentais humanos e *meta-humanos*. Da mesma forma que a Cybernética cuida da *informação* como o elemento do que permutamos com o mundo exterior, ao ajustar-nos a ele e fazermos com que nosso *ajustamento* seja nele percebido (Weiner, 1948), o Cyberdireito é forma *comunicacional* necessariamente *autorregulatória* e *regulatória*, constituída por esquemas *informacionais*, e sua empreitada atual

é entender os novos arranjos destes esquemas, os quais são integrados por entidades *intangíveis* e *desterritorializadas* em constante interdependência e reciprocidade, algo bem diverso do legado *territorializante* do Direito Romano.

CONCLUSÃO

Nosso propósito, neste modesto ensaio, foi compartilhar algumas reflexões de ordem filosófica como esforço hermenêutico sobre fenômenos atuais que impactam, de forma indelével, a existência humana. Acreditamos ser possível aprofundar nos temas aqui abordados e outros a eles conectos como projeto epistemológico de uma filosofia algorítmica orientada a meandros da *cybercultura*. Fica em aberto a indagação lançada no início deste texto: será que as formas clássicas de reflexão já não oferecem qualquer *clincher* no enfrentamento de questões (e direitos delas decorrentes) surgidas a partir de uma *revolução tecnológica* que, em poucas décadas, vem moldando a forma *de estar dos humanos no mundo*?

As pesquisas às quais venho me dedicando por mais de vinte cinco anos nas áreas de Ética e Hermenêutica com o olhar firme no horizonte da Ciência do Direito, e de certa forma ancoradas no *Idealismo Alemão*, conduziram-me inevitavelmente a me ocupar desses temas de fronteira que fazem convergir reflexões da Filosofia Prática (Ética) e da Hermenêutica Filosófica sobre os novos fenômenos jurídicos que se descortinam como *novos direitos fundamentais*. O vanguardismo e a complexidade destes temas exigem que coloquemos sob suspeita determinadas pré-compreensões estabilizadas e mantidas intocadas por séculos na história da Filosofia e do Direito da tradição ocidental. Com minha formação nas duas áreas, seria indesculpável fechar os olhos para formas comportamentais que passaram a fazer parte do cotidiano dos humanos, de modo que o benefício da dúvida sobre os caminhos epistemológicos propostos até aqui é uma exigência, não um devaneio voluntarioso que se restringe à contemplação estético-intelectiva.

Não há espaço para mais delongas argumentativas em sede de conclusão, cabendo, por derradeiro, registrar que o propósito deste artigo bastante introdutório, cujo amálgama que conecta os temas opcionalmente trazidos é a dilaceração do *conceito de humano* pelas experiências vividas pós revolução tecnológica, é noticiar algumas reflexões de um projeto de pesquisa mais amplo e que envolve as bases da Filosofia da Técnica (desde

Ernst Kapp) até a Filosofia da Tecnologia atual (com Bunge e Feenberg, por exemplo). De modo que maiores aprofundamentos serão disponibilizados em outros ensaios, ressaltando que este é um trabalho temático-conceptual, e não historiográfico. O noticiamento sobre a urgência de maiores incursões pela filosofia (da técnica e da tecnologia, inclusive) para pensar os rumos do Direito neste limiar de século parece estar longe das temáticas estafadas em vários ensaios jurídicos superficiais, mais deslumbrados com *jurimetria* do que com os rumos da nova conformação dos direitos humanos. Nesse sentido, a proposta ora apresentada a este inédito (e urgente) dossiê temático da *Revista de Direito Público*, o qual adensa os recortes *epistemológico* e *ético* no tratamento da inteligência artificial, parece trazer alguma estatura aderente ao seu desígnio. Os temas aqui costurados não nos parecem assunto trivial e desconsiderável, seja sob o ponto de vista de sua atualidade, seja sob o aspecto da carga histórica e conjuntural que ele carrega, o que, evidente, ainda merece muitas críticas cáusticas construtivas.

Em termos de historicidade, registro a intuição de que ainda não conseguimos diferençar inteligências humanas e inteligências maquínicas superpotentes senão pela capacidade misteriosa que os humanos possuem de formular *senso de aprovação e de reprovação* sobre as boas e más ações praticadas face a seus semelhantes, qualidade esta que ainda não foi replicada em máquina alguma. Ainda habitamos um mundo onde, malgrado esteja abarrotado (de) e deslumbrado (com) um cardápio *tech* sedutor, podemos considerar crível a existência de uma *alma* e de uma *consciência* exclusivamente humanas. Esta constatação *de per se* convoca (ao menos até este momento) a velha Ética, desde Platão e Aristóteles, e que sobrevive em sistemas filosóficos hoje em dia (como o de Henrique Cláudio de Lima Vaz), a tentar unir algumas pontas que ainda estão soltas quanto aos rumos do *novo* humano (pós-humano, trans-humano?). Discutir a *tecnologização* da vida exige o resgate da sua *eticização*, o que parece ser o único socorro do qual podemos dispor como um potente arsenal reflexivo a nos habilitar viver num futuro próximo e prospectar quais os passos mais acertados que conseguiremos dar nesta era *pontocom*, a qual se tornou nossa morada em mais um episódio insólito da epopeia que tem sido a história do *homo sapiens*. Isso será sempre possível enquanto ele ainda puder crer que sua *sapiência* é mais que inteligência que sabe, mas, sim, inteligência que *se* sabe, e, portanto, saber comprometido com a *sua humanidade em-si* e necessariamente *no-outro*.

REFERÊNCIAS

AGAZZI, Evandro. El impacto epistemológico de la tecnología. Argumentos de razón técnica. *Revista Española de Ciencia, Tecnología y Sociedad, y Filosofía de la Tecnología*, n. 1, p. 17-32, 1998. Disponível em: http://www.argumentos.us.es/numero1/agazzi.htm#N_1_. Acesso em: 22 jun. 2021.

_____. Conversazione con Evandro Agazzi. Interviste a Mario Alai. APhEx. Portale Italiano di Filosofia Analitica. *Giornale di Filosofia*, 21.06.2012. Disponível em: http://www.aphex.it/index.php?Interviste=557D030122020875577207020273517 17D. Acesso em: 22 jun. 2021.

AMPÈRE, André Marie. *Essai sur la philosophie des sciences ou exposition analytique d'une classification naturelle de toutes les connaissances humaines*. Paris, 1834.

BERNS, Thomas; ROUVROY, Antoinette. Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation? In: *Politique des algorithmes*. Les métriques du web. Paris: Editions La Decouverte, n. 177, 2013.

BOSTROM, N. A history of transhumanist thought. *Journal of Evolution and Technology*, v. 14, n. 1, p. 1-25, 2005.

BROCHADO, Mariah. *Ética e direito*. Pelas trilhas de Padre Vaz. Curitiba: CRV, 2021.

CAMGUIHELM, Georges. Maschine und organismus. In: *Die Erkenntnis des Lebens*. Berlin: August, 2009.

CARRAZZA, Roque Antônio. Aplicações da cibernética ao direito em outras nações (experiências resultados. opinião dos juristas). In: *Revista Justitia*, v. 36, n. 84, jan. p. 55–76, 1974.

DELEUZE, Gilles; GATTARI, Félix. *O anti-édipo*: capitalismo e esquizofrenia. São Paulo: Editora 34, 2011.

DIETRICH, Eric; HARDCASTLE, Valerie. *Sisyphus's boulder*: consciousness and the limits of the knowable. Amsterdam: John Benjamins, 2004.

DILTHEY, Wilhelm Christian Ludwig. *Einleitung in die Geisteswissenschaften* (1883). Disponível em: https://docplayer.org/46403413-Wilhelm-dilthey-1883-zum-begriff-der-geisteswissenschaften.html. Acesso em: 3 abr. 2021.

DRÖSSER, Christoph. *Total berechenbar?* Wenn Algorithmen für uns entscheiden. München: Hanser, 2016.

FACELI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. L. F. de. *Inteligência artificial*: uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, 2011.

FERRAZ JÚNIOR, Tércio Sampaio. *A ciência do Direito*. São Paulo: Atlas, 1980.

_____. *Introdução ao estudo do Direito*: técnica, decisão, dominação. São Paulo: Atlas, 2003.

GADAMER, Hans-Georg. *Verdad y método I*. 2. ed. Salamanca: Ediciones Sígueme, 1994.

HARARI, Yuval Noah. *Homo deus*: uma breve história do amanhã. São Paulo: Companhia das Letras, 2016.

HARAWAY, D.; KUNZRU, H.; TADEU, T. *Antropologia do ciborgue*: as vertigens do pós-humano. Organização e tradução de Tomaz Tadeu. 2. ed. Belo Horizonte: Autêntica, 2009.

HILBERT, Martin. A maioria dos indivíduos da espécie humana confia sua vida à inteligência artificial, todos os dias. Entrevista concedida a María José López Pourailly. *Redclara*, 29 de maio de 2018. Disponível em: https://www.redclara.net/index.php/pt/noticiasyeventos/noticias/1851-martin-hilbert-a-maioria-dos-individuos-da-especie-humana-confia-sua-vida-a-inteligencia-artificial-todos-os-dias. Acesso em: 27 jun. 2021.

HOFFMANN-RIEM, Wolfgang. *Innovation und Recht* – Recht und Innovation. Recht im Ensemble seiner Kontexte. Tübingen: Mohr Siebeck, 2016.

_____. *Teoria geral do direito digital*. Transformação digital, desafios para o Direito. São Paulo: Forense, 2021.

KANT, Immanuel. *Crítica da faculdade do juízo*. Rio de Janeiro: Forense Universitária, 1993.

KAPP, Ernst. *Grundlinien einer Philosophie der Technik*. Zur Entstehungsgeschichte der Kultur aus neuen Gesichtspunkten (1877). Hamburg: Felix Meiner Verlag, 2015.

KELSEN, Hans. *A justiça e o direito natural*. 2. ed. Coimbra: Arménio Amado, 1979.

LATOUR, Bruno. *Jamais fomos modernos*. São Paulo: Editora 34, 2013.

LAZZARATO Maurizio. *Signos, máquinas, subjetividades*. São Paulo: Edições SESC SP, 2014.

LÉVY, Pierre. *Cibercultura*. São Paulo: Editora 34, 2010.

MORA, José Ferrater. *Dicionário de Filosofia*. São Paulo: Edições Loyola, t. I e IV, 2000.

MORE, M. *Transhumanism*: toward a futurist Philosophy, 1990. Disponível em: https://www.ildodopensiero.it/wp-content/uploads/2019/03/max-more-transhumanism-towards-a-futurist-philosophy.pdf. Acesso em: 20 jun. 2021.

MOROZOV, Evgeny. *Big tech*: a ascensão dos dados e a morte da política. São Paulo: Ubu, 2018.

RICOEUR, Paul. *Le conflit des interprétations*. Essais d' Hermenéutique I. Paris: Editions du Seuil, 1969.

RODRIGUES, Bruno Alves. *A inteligência artificial no Poder Judiciário*: e a convergência com a consciência humana para a efetividade da justiça. São Paulo: Thomson Reuters Brasil, 2021.

SALGADO, Joaquim Carlos. *A ideia de justiça no mundo contemporâneo*: a interpretação e aplicação do Direito como *maximum* ético. Belo Horizonte: Del Rey, 2006.

SCHLEIERMACHER, F. D. E. (1826). *Hermeneutik und Kritik*. Hg. v. Manfred Frank. Frankfurt/M.: Suhrkamp, 1977.

TOEPFER, Georg. Der Begriff des Lebens. In: KROHS, Ulrich; TOEPFER, Georg (Hrsg.). *Philosophie der Biologie*. Eine Einführung. Frankfurt: Suhrkamp, 2006.

TUTT, Andrew. An FDA for algorithms. In: *Administrative Law Review*, 2017. Disponível em: http://www.administrativelawreview.org/wp-content/uploads/2019/09/69-1-Andrew-Tutt.pdf. Acesso em: 14 mar. 2021.

VAIHINGER, Hans. (1911). *Die Philosophie des Als Ob*. Paderborn: Salzwasser-Verlag, 2013.

VAZ, Henrique Cláudio de Lima. *Escritos de filosofia II*: ética e cultura. São Paulo: Edições Loyola, 1993.

_____. *Escritos de filosofia V*: introdução à ética filosófica 2. São Paulo: Edições Loyola, 2000.

WIENER, N. (1948). *Cybernetics*: or control and communication in the animal and the machine. N.Y: The Technology Press, 1961.

WIENER, Norbert. *The human use for humains beings*. Cybernetics and society. London: Free Association Books, 1989.

YOUNG, S. *Designer evolution*: a transhumanist manifesto. New York: Prometheus Books, 2005.

ZIVIANI, Nívio. A quarta revolução tecnológica. Computação cognitiva e a humanização das máquinas. Disponível em: https://homepages.dcc.ufmg.br/~nivio/papers/a-quarta-revolucao-industrial-fonte-julho2017.pdf. Acesso em: 26 jun. 2021.

_____. *Projeto de algoritmos com implementações em Pascal e C*. São Paulo: Editora Unicamp, 1999.

Mariah Brochado | *E-mail:*mbrochado@gmail.com

Pós-Doutorado em Filosofia pela Philosophisches Fakultät/Ruprecht-Karls-Universität Heidelberg – Alemanha. Doutorado, Mestrado e Especialização em Filosofia do Direito pela Faculdade de Direito da Universidade Federal de Minas Gerais – UFMG. Professora Associada da Graduação e da Pós-Graduação da Faculdade de Direito da UFMG. Coordenadora do Projeto Cyberdireito e Filosofia Algorítmica (FDUFMG). Presidente da Comissão de Inteligência Artificial no Direito da OAB-MG. Associada da *The Society for Philosophy and Technology* (SPT) – EUA. Associada Honorária da União Ibero-Americana de Juízes. Foi Secretária de Estado de Casa Civil e de Relações Institucionais de Minas Gerais (2015-2018) e Presidente da Comissão que elaborou o Decreto-regulamentar da Ciência, Tecnologia & Inovação em Minas Gerais (Decreto 47.442, de 04 de julho de 2018). Pesquisadora em Filosofia e Teoria do Direito, com ênfase em Hermenêutica, Ética e Filosofia da Tecnologia aplicada ao Direito.

Data de submissão: 27 de setembro de 2021.
Data de aceite: 10 de janeiro de 2022.

What is it Like To Be an Artificial Intelligence? Nagel's View From Nowhere and Artificial Intelligence as Subject of Law

Como É Ser uma Inteligência Artificial? A Visão a Partir de Lugar Nenhum de Nagel e a Inteligência Artificial como Sujeito de Direito

STHÉFANO BRUNO SANTOS DIVINO[1]
Centro Universitário de Lavras (UNILAVRAS).

RODRIGO ALMEIDA MAGALHÃES[2]
Pontifícia Universidade Católica de Minas Gerais (PUC/MG).

ABSTRACT: This paper has as its research problem the following question: what is it like to be an artificial intelligence? It aims to critically analyze the epistemological and semantic aspects developed by Thomas Nagel in What is it like to be a bat and The View from Nowhere, demonstrating the relationship between physicalism and subjectivity and its application to artificially intelligent beings. We chose to approach these two works because of the author's importance in analytical philosophy and the approach to consciousness. The analysis shows that the defense of artificial intelligence as a subject of law is intrinsically based on physicalism. However, in refuting it, Nagel does not offer an alternative outside the scope of dualism. Thus, the Procedural Theory of the Subject of Law is developed with stages of emancipation of the being against the law. As a result, it is verified that the reductive physicalist vision is insufficient to substantiate the condition of the subject of law of an artificial intelligence as a legal and political being in the social order. However, if the three stages of its formation (emancipation, interspecies recognition, and personification) are observed, the possibility of achieving the condition under analysis is assumed. It is concluded that it is unverifiable to know what it is like to be an artificial intelligence. In the current scientific stage, an artificially intelligent being cannot (yet) be considered a subject of law, under penalty of characterization of instrumentalism. The methodology of integrated, analytical, deductive, and bibliographic research is used to obtain these results and conclusions.

KEYWORDS: Artificial intelligence; subject of law; Procedural Theory.

1 Orcid: http://orcid.org/0000-0002-9037-0405.
2 Orcid: http://orcid.org/0000-0002-8869-2114.

RESUMO: Este trabalho tem como problema de pesquisa o seguinte questionamento: como é ser uma inteligência artificial? Objetiva-se analisar criticamente os aspectos epistemológicos e semânticos desenvolvidos por Thomas Nagel em *What is it like to be a bat* e *The view from nowhere,* demonstrando a relação entre o fisicalismo e a subjetividade, bem como sua aplicação nos entes inteligentes artificialmente. Opta-se pela abordagem nessas duas obras pela importância do autor no âmbito da filosofia analítica e na abordagem da consciência. A análise demonstra que a defesa da inteligência artificial como sujeito de direito está intrinsecamente pautada no fisicalismo. Contudo, ao refutá-lo, Nagel não oferece uma alternativa fora do escopo do dualismo. Dessa forma, desenvolve-se Teoria Procedimental do Sujeito de Direito com seus estágios de emancipação do ser perante a ordem jurídica. Como resultado, verifica-se que a visão redutiva fisicalista é insuficiente para fundamentar a condição de sujeito de direito de uma inteligência artificial enquanto ser jurídico e político na ordem social. Contudo, caso sejam observados os três estágios de sua formação (emancipação; reconhecimento interespécie; e personificação) assume-se a possibilidade de conquista da condição em análise. Conclui-se que é inverificável saber como é ser uma inteligência artificial e que no estágio científico atual um ente inteligente artificialmente (ainda) não pode ser considerado sujeito de direito, sob pena de caracterização do instrumentalismo. Para obtenção desses resultados e conclusões utiliza-se a metodologia de pesquisa integrada, analítica, dedutiva e a técnica de pesquisa bibliográfica.

PALAVRAS-CHAVE: Inteligência artificial; sujeito de direito; Teoria Procedimental.

1 INTRODUCTION

With the development of artificial intelligence, questions ranging from responsibility, data treatment, discrimination, or acting as an autonomous agent have been addressed in the legal, social, and philosophical scenarios. However, in the face of the novelty and unpredictability in the actions of these entities, the discourse on ethical limits and their responsibility has been opaque. Thus, a more refined and objective discourse is needed to clarify this problematic issue.

Given the myriad of facts and cases to be addressed, this article chooses to focus on the epistemological aspect of Artificial Intelligence. With this approach, we intend to demonstrate the differences and similarities between the cognitive capacity of an AI and a human being. To this end, the research problem of this essay is: what is it like to be artificial intelligence? The question is an analogy to Nagel's seminal essay "What is it like to be a bat?". In this article, Nagel discusses the relationship between different viewpoints and how subjectivity and objectivity must be correlated through a critique of physicalism.

Developing this analysis is essential since it is assumed that Law is a science formed by subjective criteria in an objective-looking world. However, it must be verified that an objective world must hold different subjective views. To this end, the view from nowhere developed by Nagel is used to describe the situation, identification, and participation of the subjective in the objective, correlating the mind-body problematic, will, and the different world views to the social and legal complexity. It aims to problematize: what is consciousness? What is the self (in the world)? What position does the individual's thought occupy in an objective world? Is there a way to reconcile the subjective, first-person view with the objective, third--person view?

The result is that different points of view can and should be incorporated into an objective reality. Besides, it is verified that for Law to be complete, and anthropomorphism must be left aside from the moment that some situations and beings cannot be understood by humanity. Therefore, although artificial intelligence can be a subject of Law[3], its characterization becomes unfeasible since the current technological, scientific stage does not allow these beings to act in a genuinely autonomous manner to the point of demonstrating their different points of view. From this approach, recognizing artificial intelligence as a subject of Law necessarily goes through instrumentalism.

However, these considerations cannot be easily extracted from Nagel's approach, for the author, by denying physicalism, does not present a viable solution for the insertion of different worldviews, especially in Law. At this point, the procedural theory of the subject of Law is inserted (Divino, 2020), demonstrating the stages of development of the being before the legal order and against its power. First, one must keep in mind the understanding of the subject of Law and differentiate it from the term person to verify how and where it is formed. It is evident that it is a constant struggle against the legal norm itself, a relationship traced back to Foucault. However, Nagel's nowhere vision offers alternatives to avoid the essentially subjective formation of Law. In this way, artificial intelligence could be considered

3 "A subject of law is one's who can exercise his rights and duties in the legal system without someone's representing him. It is conquering his position in the system. It is winning a battle against the domination process. It is being rational and linguistic subject to contribute to society for their interspecies relations. It is an active person in the legal system who can do whatever he wants since the law does not prohibit it. This structure will be developed now. It is not a definitive view (so far). It's a dialogue between ideas for the law system evolution" (Divino, 2020, p. 184).

a legal subject when it possesses sufficient autonomy and manifestations similar to human beings.

However, the formation of the subject of Law must also go through interspecies recognition. This recognition is made from the moment that Law is considered as a product of language. The syntactic, semantic, and epistemic aspects are united to produce a coherent result and allow human society to give up its anthropomorphic egoism and recognize artificially intelligent beings as autonomous political and legal beings.

Given the above, we conclude that it is unintelligible to understand what it is like to be artificial intelligence. Therefore, it is also unintelligible to attribute an autonomous regulation to these entities in the current scenario due to the lack of scientific and technological progress. However, this does not mean that these entities can be left aside and their worldviews denied by the possible absence of subjectivity. With the procedural theory of the subject of Law, artificial intelligence can be elevated to a social and legal level similar to the human one, as long as it demonstrates this understanding to deserve its legal protection. Integrated, analytical, deductive, and bibliographic research is used to obtain these results and conclusions.

2 WHAT IS IT LIKE TO BE AN ARTIFICIAL INTELLIGENCE? HUMAN AND NON-HUMAN COGNITIVE CAPABILITIES FROM NOWHERE VIEW

The discussion between body and mind began with Cartesian dualism and its development with analytical philosophy in the 20th century. However, the actuality of this approach remains when one takes a comparative look at the consciousness of human beings and non-human beings. In the latter case, the term "non-human" should also be understood as non-biological. To know whether an artificial intelligence deserves autonomous treatment and to possess rights and duties, we need to understand what it is like to be an Artificial Intelligence[4]. Therefore, this section aims to demonstrate how an artificially intelligent entity should be treated according to your point of view.

4 McCarthy says (2007, p. 2-15): "It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable".

According to Russell and Norvig (2010, p. VIII), AI can be defined "[...] as the study of agents that receive percepts from the environment and perform actions".

This starting point becomes necessary to justify the reasons and rationale for recognizing rights and duties for a being that does not fit as a human being. Why should we recognize the rights and responsibilities of artificial intelligence? How can artificial intelligence modify reality in its objective aspect? Is it possible to talk about the experience without tying it to objectivity? These are the questions that will guide this section and help answer the proposed problem.

First, it should be recognized that the construction of Law is something essentially subjective. In other words, Law is formed as a result of cultural practices in a given society. Therefore, what may be accepted and recognized in one country may not be so in another. Thus, the purpose and result of legal production are to respond to the interests of society, and this production occurs essentially through the mental phenomena of political beings. Therefore, the accurate result is linked to solving a given country's political or social issues.

One of the products of this legal, social, and political practice is the concretization of the subject of Law. However, this concretization is only possible through experiences, which are understood and interpreted semantically and epistemologically through consciousness. This is where Nagel can contribute to the approach. First, imagine what it is like to be a person without fundamental rights. In this case, there is the possibility of imagination and even proof of this fact. You can think, for example, of slaves, where fundamental rights and guarantees were ignored. So, you can imagine yourself as a person or a human being without fundamental rights and guarantees to the extent that you are a human being with fundamental rights and guarantees.

However, what is it like to be a freezer? What is it like to be a rock? What is it like to be a non-life object? You cannot feel or imagine because there is no previous experience that allows you to feel that. To be an inanimate object is not to be something static in darkness without any correspondence from its environment. It can be said to be so because inanimate and lifeless objects do not experience moments. There is no subjective relationship in a rock experiencing darkness. If you close your eyes, you can feel and experience darkness. However, if you close your eyes again, you cannot handle the experience as a rock. Inanimate objects, therefore, have no experience at all. There is nothing in our sense we can match to feel in this way. What allows experience is consciousness. What

it is like to be that being is only possible throughout by experience and consciousness.

Nevertheless, where does the artificially intelligent entity fit into this scenario? If a developer creates an artificial intelligence that can detect colors or numbers, his work will be successful if the programming is successful. Therefore, both a human being and artificial intelligence would see colors and numbers from their observation. However, there is a difference in the point of view of these beings. In the case of the human being, it can be said that consciousness makes it possible to experience analysis. Seeing a green or red object is a physically objective analysis and a subjective experience of that reality. In artificial intelligence or an electronic device designed to detect color or number, there is no relation of experience because of the inexistence of the mind in that being. They cannot see what the color or the number looks like. Because they are not conscious, but humans are.

However, there is a point to consider. Not all mental states are conscious. When we are asleep, mental states continue to happen, but we are not aware of this. Beliefs are also good examples of intrinsic mental conditions in the mind but are not always conscious. In this situation, a person may firmly believe that the whale is the largest mammal in existence and not always be aware of this fact.

Furthermore, there are situations that, although experienced, consciousness is mitigated into a background. Just think about the piece of clothing you are wearing right now. You may have spent a great deal of time wearing it and experiencing it in an unconscious mental state, but from the moment you read this passage, you have consciously directed your attention to feeling this garment.

In short, "an organism has conscious mental states if and only if there is something that it is like to be that organism-it is like the something organism" (Nagel, 1974, p. 436). However, why Nagel's vision is essential here? Now we can talk about the bats. Bats have an echolocating system that humans do not. We can smell, taste, touch, see and hear, but we do not have an echolocating system. So, Nagel does the question, "What is it like to be a bat?". What is it like to fly and map all the environment around you through an echolocating system? Can we understand what it is like for a bat to echolocate? For Nagel, the answer is negative because humans have no idea how it feels like. So, we cannot experience something that we do not have and cannot understand. However, what is the point of this example?

Nagel said: "Whatever may be the status of facts about what it is like to be a human being, or a bat, or a Martian, these appear to be facts that embody a particular point of view" (Nagel 1974, p. 441).

In other words, Nagel is saying that there are facts that can only be known by getting into a particular/subjective perspective[5]. And why is this observation important? Because as Law must, or at least should, be based on reality, its construction from thought must have as its objective the regulation of those who construct it. In other words, the formation of Law is essentially subjective. It exists as a function of social code towards its creators. Thus, what are the interests of artificial intelligence to recognize it as an autonomous entity in the contemporary scientific and legal scenario? The answer to this question is impossible to verify at this point. Since artificial intelligence is developed under a strictly objective aspect, it is inferred that there are no subjective aspects in its programming. Therefore, the interests expressed and manifested by these entities seem to be nothing more and nothing less than simulations of the interests of their developers.

As much as there are findings on the development of machine learning and deep learning techniques that supposedly authorize granting a certain degree of autonomy to artificial intelligence, one should be cautious about affirming the existence of mental states in these beings. Furthermore, why is the fact of mental states important in this case? Because it makes no sense to insert an artificially intelligent entity into the legal order that cannot act autonomously and is incompatible with the creation of the legal system itself. This would be a fundamental instrumentalist approach.

At this point, Nagel's argument becomes more acceptable: in refuting physicalism, a reductionist theory that defends the idea that aspects of the world are reduced to physical justifications. In this sense, according to physicalism, one could say that rain, lightning, or a storm could be analyzed under the objective aspect and reduced to physical explanations about their formation. Lightning, for example, would be an electric discharge that moves atoms. The modification of water states could understand rain. Moreover, a storm could be explained by the relationship between the various climatic factors and air densities. Therefore, physicalism starts from the assumption

5 "The more different from oneself the other experiencer is, the less success one can expect with this enterprise. In our own case we occupy the relevant point of view, but we will have as much difficulty understanding our own experience properly if we approach it from another point of view as we would if we tried to understand the experience of another species without taking up its point of view" (NAGEL, 1974, p. 442).

that the relations existing in the world are essentially objective and reduced to physical explanations. Physicalism, therefore, cannot justify or explain the experience because it gets away from any subjectivity.

When we use physicalism to approach the relationship between body and mind, the premise that we can find is that the mind would be an exclusively physical phenomenon arising from the chemical relationship existing in the brain. In this sense, when the object of study is biological beings, brains have minds and, consequently, can think through their consciousness. However, when we talk about non-biological beings, this idea is not falsified. For physicalism, an artificially intelligent being could be endowed with consciousness to the extent that it has physical relations in its hardware and software. In this way, artificial intelligence would think and present its objective view under its worldview.

However, physicalism denies the subjective part existing in the world. Understanding what rain or a storm is from an objective point of view is not the same as experiencing it. Therefore, experience is linked to the subjective aspect of a worldview, which will be responsible for forming consciousness.

When we talk about Law, its formation is essentially objective and based on different points of view. Hard sciences cannot explain how Law works because Law is made from subjective points of view. If we consider it accurate, the subject of Law must be a being with a subjective point of view. In this sense, how can we recognize that artificial intelligence possesses subjectivity if we do not yet understand how the mind works? This is a question that is impossible to answer at the moment and, at worst, will not be answered because there are things that human beings cannot explain.

So, what is it like to be artificial intelligence? We do not know. There is no subjective perspective to understand how an artificial intelligence sees and experiences the world without enough technological resources. So far, the argument or thesis for granting autonomy to artificial intelligence in Law is essentially physicalist, reductionist, and instrumentalist. It is a good way for developers and companies to avoid liability from the damages caused by their actions.

Nevertheless, we might have something in Nagel's argument to explain how artificial intelligence can be part of the world. Nagel proposes that one should project X into the world as a thing that interacts with others and question it as the world must be when viewed from no particular point of view, to present X as it presents itself from its point of view (Nagel, 1986,

p. 66) In other words, X should think of itself as part of the world and see itself from the outside. Nagel exemplifies:

> Suppose all the nerves feeding sensory data to my brain were cut but I were somehow kept breathing and nourished and conscious. And suppose auditory and visual experiences could be produced in me not by sound and light but by direct stimulation of the nerves, so that I could be fed information in words and images about what was going on in the world, what other people saw and heard, and so forth. Then I would have a conception of the world without having any perspective on it. Even if I pictured it to myself I would not be viewing it from where I was. It might even be said that in the sense in which I am now TN, I would under those circumstances not be anyone (Nagel, 1986, p. 66).

It is inferred that the objective self is only part of a person's point of view, and their objectivity develops to different degrees at different stages of life and civilization. There is the possibility of convergence of other people's world views without a center, so there is a close relationship between objectivity and intersubjectivity. The quest is to share a conception, a point of view, between X and the other individuals in the world. This is why the objective self of X is not singular, and each Y possesses one. For Nagel, the objective self is not a distinct entity; each individual, besides being an ordinary person, is a particular objective self, a subject of a conception of reality devoid of perspective[6].

It should be noted that, although the impersonal conception of the world does not grant any particular position to individual X, it is tied to his perspective and develops from it. This does not mean that the world is the world of X: he is not its subject, but only one of the persons contained in it, and none of them is its center or focal point. In this way, X is the logical focus of an objective conception of the world and a particular being that does not occupy any central position. Since X possesses or is an objective self, one can express meaningful identity by alluding to X indicatively as a self and, objectively, to the person publicly identifiable as X, and also make both references from the same point of view as someone who possesses the

6 "The objective self should be able to deal with experiences from any point of view. It in fact receives those of TN directly, but it treats on an equal footing those it receives directly and those others it learns about only indirectly. So far as its essential nature is concerned, it could base its view of the world on a different set of experiences from those of TN, or even none at all coming directly from a perspective within the world, for in itself it has no such perspective. It is the perspectiveless subject that constructs a centerless conception of the world by casting all perspectives into the content of that world" (Nagel, 1986, p. 67)

same objective conception of the world that contains X. Since the objective concept has a subject, what allows the objective and subjective views to be joined is the fact that there is the possibility that he is present in the world. The purely objective conception, without the inclusion of the subjective, rejecting their union, will leave out something authentic and remarkable (Nagel, 1986, p. 67).

For Nagel (1986), the objective self must be the reference in virtue of something greater whose inclusion in the world is not apparent. The only meaningful aspect under which the individual can refer to himself subjectively, provided only by the objective conception of the world because he is the subject of that conception. Nagel (1986 p. 68) describes the degree of autonomy of the objective self such that it would have sufficient independence from the private self to have a life of its own. Sometimes the author treats it as if it were a distinct part of the mind, but this does not mean that it should be given a metaphysical nature incompatible with Cartesian dualistic theory. The problem of reconciling the subjective, first--person view with the objective, third-person perspective finds an answer in these premises[7].

Going out of ourselves and seeing the world from nowhere within it is a means for expanding knowledge and our doubts about ourselves and the world, which never end. The problem arises from the premise: "how limited beings like ourselves can alter their conception of the world so that it is no longer just the view from where they are but in a sense a view from nowhere" (Nagel 2004, p. 73). This, in turn, must include and understand the fact that the world contains beings who have such a view, explaining why the world appears to them to be as it is before they form that conception and describing how they can arrive at that conception. The particular viewpoint would be only instrumental, not essential (Divino, 2021, p. 249). It would appear that the most objective view that can be achieved must rest on a subjective basis and free examination and that one can no longer abandon the particular point of view but only change it. Reductive and heroic skeptical theories of knowledge present answers to this problem. "Skeptical possibilities are those according to which the world is completely different from how it appears

7 "We can account for the content of the philosophical thought 'T am TN' if we understand the 'I' as referring to me qua subject of the impersonal conception of the world which contains TN. The reference is still essentially indexical, and cannot be eliminated in favor of an objective description, but the thought avoids triviality because it depends on the fact that this impersonal conception of the world, though it accords no special position to TN, is attached to and developed from the perspective of TN" (Nagel, 1986, p. 68).

to us, and there is no way to detect this" (Nagel 1986, p. 74). Reductive theories advocate a view of the world not as it is but as it presents itself to us. "Assuming that we do know certain things, and acknowledging that we could not know them if the gap between content and grounds were as great as the skeptic thinks it is, the reductionist reinterprets the content of our beliefs about the world so that they claim less" (Nagel, 1986, p. 72). In turn, recognizing the large gap between the foundations of our beliefs about the world and the contents of those beliefs, heroic theories subject this gap to a realist interpretation in an attempt to bridge it without narrowing it.

Nagel refutes skepticism, defending the idea that it would be a means of recognizing our situation, but not a way of coveting us to go on and pursue something like knowledge because our natural realism prevents us from settling for a purely subjective view. The problem with heroic theories, on the other hand, lies in epistemology (the first-person problem of what to believe and how to justify one's beliefs), which need not necessarily be reductionist and may be vulnerable to metaphysical arguments just like Descartes' dualist theory. Reductionist theories, by reinterpreting the content of our beliefs about the world as it presents itself to us to reduce its claims, naturally cannot escape skepticism because constructing a reductionist analysis of claims about the world that has a minimum of plausibility without leaving gaps between the grounds and the content is challenging (Nagel, 1986).

Nagel, therefore, leans toward a rationalist theory. "The conditions of objectivity that I have been defending lead to the conclusion that the basis of most real knowledge must be a priori and drawn from within ourselves" (Nagel, 1986, p. 85). For Nagel, the role played by the particular experience and the action that the world exerts in the specific through its perspectives can only be selective. Hence,

> If the possibilities, or at least some of them, are available a priori to any mind of sufficient complexity, and if the general properties of reality are fairly uniform throughout, then the pursuit of objective knowledge can be expected to lead to gradual convergence from different starting points. (Nagel, 1986, p. 86).

However, Nagel (1986, p. 86) points out that this limit of convergence is a consequence of the relationship between reality and mind, not the definition of truth. For this reason, the author (Nagel 1986, p. 87) believes that mental capacities do not entirely mirror reality but assumes that all

individuals potentially have in their heads the possibilities that will be revealed in the course of the next millennium by scientific advances. His basis, therefore, falls within the rationalist tradition of an anti-empiricist theory of knowledge, for the author presupposes an unheard-of property of the natural order that we are not aware of, such as the role that Descartes attributed to God.

Nagel commits to realism for the defense of his position. This theory brings the conception that the world is independent of our minds. Nagel defends a form of realism "according to which our grasp on the world is limited not only in respect of what we can know but also in respect of what we can conceive. In a very strong sense, the world extends beyond the reach of our minds" (Nagel, 1986, p. 92). In taking this stance, Nagel (1986) abandons idealism[8]. This theoretical axis holds that what exists is what we can think or conceive, or what our descendants or we might come to believe or imagine, and idealistic conceptions, which hold reality as correlative to the mind in a much broader sense, that is, it includes infinite minds. The realism defended by the author says that the world is perhaps inconceivable to our minds since it would be independent of our possible representations and capable of extending beyond them (Nagel, 1986, p. 93). The incompleteness of objectivity is taken into consideration, for in some aspects, it does not correspond to reality and is not always the best method of knowledge. Therefore, given human nature, what exists and we can think they are two different things, and the latter may be smaller than the former (Nagel, 1986, p. 93)[9]. Thus, Nagel (1986, p. 96) believes that "reality extends

8 Berkley (1710, p. 2 and 23) defended the thesis that "for unthinking things, to exist is to be perceived; so they couldn't possibly exist out of the minds or thinking things that perceive them". For him, this is evident when trying to conceive a non-perceived object because it is impossible. If we think of a book on a shelf unperceived, there is an evocation of the perceptual image, and thus it is not something unperceived. "But", you say, "surely there is nothing easier than to imagine trees in a park, for instance, or books on a shelf, with nobody there to perceive them". I reply that this is indeed easy to imagine; but let us look into what happens when you imagine it. You form in your mind certain ideas that you call "books" and "trees", and at the same time you omit to form the idea of anyone who might perceive them. But while you are doing this, you perceive or think of them! So your thought- experiment misses the point; it shows only that you have the power of imagining or forming ideas in your mind; but it doesn't show that you can conceive it possible for the objects of your thought to exist outside the mind. To show that, you would have to conceive them existing unconceived or unthought-of, which is an obvious contradiction. However hard we try to conceive the existence of external bodies, all we achieve is to contemplate our own ideas. The mind is misled into thinking that it can and does conceive bodies existing outside the mind or unthought-of because it pays no attention to itself, and so doesn't notice that it contains or thinks of the things that it conceives. Think about it a little and you will see that what I am saying is plainly true; there is really no need for any of the other disproofs of the existence of material substance.

9 "It has already been argued that in various respects the pursuit of objectivity can be carried to excess, that it can lead away from the truth if carried out in the wrong way or with respect to the wrong subject matter. That is one way in which objectivity does not correspond to reality: it is not always the best mode of understanding. But human objectivity may fail to exhaust reality for another reason: there may be aspects of reality beyond

beyond the reach of possible human thought, since this would be closely analogous to something which is not only possibly but actually the case", because "the existence of unreachable aspects of reality is independent of their conceivability by any actual mind" (Nagel, 1986, p. 97).

Therefore, although human beings cannot answer the question "What is it like to be an artificial intelligence" one can include their point of view in the world by recognizing that the human mind is limited and does not represent the entire content of the world. However, this recognition is only possible through the procedural theory of the subject of law, which tends to present a balance[10] between objectivity and subjectivity and the refutation of instrumental solipsistic physicalism.

3 THE PROCEDURAL THEORY OF SUBJECT OF LAW: HOW CAN ARTIFICIAL INTELLIGENCE BE PART OF THE LEGAL SYSTEM?

This section aims to demonstrate how artificial intelligence can be part of the legal system. For the anxious reader, the answer is *as a subject of law*. However, to reach this result, the reasoning adopted goes through a brief historical background of the formation of the subject of Law. Furthermore, one must change the perspective that Law arises only from an exclusively subjective point of view, as proposed by Nagel. In this sense, it is not to adopt a reductionist or physicalist view as discussed above, but a realist idea that recognizes the incompleteness of the human biological mind in representing the various aspects of the world. To develop these two approaches, the idea of Law as an emancipatory instrument and Law as a tool for interspecies recognition is developed. These stages stem from the procedural theory of the subject of Law that aims to explain how it is formed and inserted into the legal order.

---

its reach because they are altogether beyond our capacity to form conceptions of the world. What there is and what we, in virtue of our nature, can think about are different things, and the latter may be smaller than the former. Certainly we are smaller, so this should not be surprising. Human objectivity may be able to grasp only part of the world, but when it is successful it should provide us with an understanding of aspects of reality whose existence is completely independent of our capacity to think about them – as independent as the existence of things we can't conceive" (Nagel, 1986, p. 93).

10 This balance is only possible by assuming that any objective perspective of view must contain a particular point of view, and this one is subjective. The procedural theory of the subject of law aims to recognize not only sentient beings as a subject but those who use rationality as well to communicate and struggle for their recognition in the society. The main argument in our theory is that any rational being (even it is not human) can be a subject and recognized as well by the Law.

First of all, explaining the subject of Law in terms of traditional theories becomes elementary. Jusnaturalism conceives the subject of Law as a natural condition of man (lato sensu). The attributions. Law, in this case, corresponds to the primary art part of a moral reality used to give the thing of each one (Hervada, 2006, p. 16 and 131), which designates and recognizes the being in his position as a natural subject of Law, a unitary composition of the socio-relational system itself. On the other side, the subject of Law appears as a purely normative determination. There is a reduction of this figure to a complex of legal norms. Both the natural person and the legal person are figurative entities of a figuratively legal reality expressed in the concept of person, whose concept is merely the personification of this unity (Divino, 2021). Ultimately, legal duties and subjective rights are enacted by legal norms and reduce the unitary problem of the person to a complex of norms (Kelsen, 1967, p. 193-194).

However, both naturalism and positivism have misconceptions. Explaining the condition of the subject of Law according to state of the art and of things should start from the differentiation between persona and subject of Law. Persona corresponded, in ancient times, to the invocation of an artificial and fictitious dimension, where the man (real being) used a mask (*prosôpon*) to personify himself and put himself in another's skin, aiming to get rid of his own (Viola, 2017, p. 14). The person was not an individual, nor was the individual a person. The notion of *in-dividuus*, the *undivided*, has its equivalent in the Middle Ages, the idea of the atom, an order, an *estamento* capable of being recognized as a social being (Martins-Costa, 2010, p. 69). *Prospon* represented having *(habere personam*) and not being. There was no talk of a symbiotic existence between person and self, nor consciousness, since this correlation has its origin only in Cartesian philosophical principles (Descartes, 2004, p. 46). Its attribution, therefore, was more linked to the political aspect and less to the legal one since the usefulness of the mask was represented as a place of speech disconnected from the being that uttered it.

On the other hand, the subject of Law develops markedly in liberal humanistic legal thought with the *fictio juris* that human beings are equal before the Law and, therefore, holders of the same rights and obligations. This situation, however, is not simple to be realized. The construction of subjectivity or subjectivation is excellently analyzed by Foucault. The author performs a critique within the legal grammar, directing his attention to a critical or practical attitude towards a re-signification. This means that the target subject of normative power uses the normative tools at his disposal

to modify them and construct himself. Foucauldian critique is the art of not being ruled, "a certain way of thinking, of saying, of acting equally, a certain relation to what exists, to what one knows, what one does, a relation to society, to culture, a relation to others as well" (Foucault, 2015, p. 31). This is because the subject of Law in Foucault serves as resistance to the normative domination system for constructing one's subjectivity.

The inertia of the subject concerning the practice of criticism leads him to governmentalization, which is defined as a social practice of individual subjection by mechanisms of power that claim a truth (Foucault, 2015, p. 35). Criticism has its importance from the moment the subject gives himself the right to "question truth about its effects of power and power over its discourses of truth" (Foucault, 2015, p. 35). To the subject is entrusted the task of fight in the politics of truth. For Foucault (1995, p. 234), the control of power can be realized by anti-authoritarian struggles that affirm the right to be different and emphasize everything that makes subjects genuinely individual. "The subject, therefore, suffers the effects of power, and it is from these effects that he can be identified and constituted as an individual" (Silva; Rodrigues, 2019, p. 2297). For Foucault (1999, p. 35), the subject "is an effect of power and is, at the same time, to the same extent that it is an effect of it, it's intermediary: power transits through the individual it has constituted".

The subject of Law emerges as a resistance and, at the same time, a concession of the being to the Law from the moment the latter exercises the power of domination over the persona and prevents it from exercising the practices of the self. Foucault's constitution of the subject of Law can be seen as a relation of domination between the legal system aimed at normalization practices, where the being reacts against them and constructs itself. There is an asymmetry between the social relations inscribed by the subject and the norm in which the former is constituted through practices contrary to the coercive attribution of a specific identity.

Thus, the sense of differentiation between persona and subject of Law only exists if the regulatory legal system does so or attributes qualities and legal situations to one and not the other. This means that if every person is a subject of rights and duties, the differentiation between non-person and non-subject of rights and responsibilities is undermined unless the legal system grants rights and obligations to the non-person.

The second thought to be raised is that Foucault seems to be correct but in part. There is a correlational duality between one who subjects himself to someone, by control and dependence, as well as the transition from subject to his autonomous synthesis, by a conscious or self-aware step, both of which, according to Foucault (1995, p. 235), "suggest a form of power that subjugates and renders subject". In this sense, what Foucault tries to avoid is reducing social plurality through the normalizing system. In other words, he tries to prevent normalization through normalization. It is up to the subject and him to pay attention to the regulatory forms of power to transform himself and control his definitive settlement. In short, the subject must integrate the Law as a claimant agent of its claims and rights. This is the Foucauldian correctness. The process of constitution of the subject of Law must be emancipatory. There must be a counter-situation of subjection and resistance to the intrinsic power to him at that moment. Recognition becomes legitimate from the moment Law recognizes the critical practice claims produced by the confrontation with dominant discourses.

To consider an artificially intelligent entity as a subject of Law seems necessary to go through this claiming process[11]. The entity must demonstrate to society and the legal system how its capacities are required and what rights it is possible to attribute before the moment of the praxis of the self. The simple concession of this legal position exalts the utilitarian character and ignores years of contributions and class struggles for social emancipation. It is up to the mechanical self to contradict the legal norm that restricts the position of rights and duties to human beings and terminates the obligation of obedience existing in this legal system. Law must act with its spectrum of legality, not normalizing and oppressing. If granted in the contemporary molds, the subjectivation of artificially intelligent beings tends to increase social complexity considerably. "By complexity, we mean that there are always more possibilities than can be realized" (Luhmann, 1983, p. 44).

11 In this context, only systems classified as ASI would be able to fulfill the requirements of this process. Regarding the ANI and AGI systems, although the European Parliament has proposed for the possibility of legal personhood not in attention to (possible) self-awareness, but as a system to regulate liability for damages in cases of more complex systems that prevent the attribution of harm to a human agent, such position should not prosper. A first argument is that personhood is only one of the (final) stages of the constitution of the subject of law. The entire emancipatory process in which self-awareness is considered essential for understanding the factual and legal aspects is indispensable for its configuration. The simple attribution of personality can work as a dodge of responsibility or just a distortion of the economic risks of the business. In a simple visualization, it can work as target instruments of conducts (desired or not) liable to civil responsibility, something very similar to what we have to the legal personality of companies and corporations, which had in its origin this justification to separate the patrimonial aspect of its constituent/exerciser of the economic activity originated from the positive or negative results arising from this social practice.

Contingency, understood as "the fact that the possibilities aimed at the other experiences could be different from those expected", referring to something deceptive, non-existent, or unattainable means the forced selection of social situations capable of generating unnecessary dangers and risks (Luhmann, 1983, p. 45). Therefore, for an artificially intelligent entity to be considered as a subject of Law, the emancipatory process must be carried out towards Law, but not only.

After the emancipation stage, the view from nowhere should be applied here. In this sense, society and Law must understand how the anthropomorphic view is limited and recognize other non-human and non-biological beings in the system. Law is represented by language, semantics, and syntactic authorizes the agent to use Wittgenstein's language games to elaborate public rules capable of legitimizing the emancipatory process. It is the Law itself that, within its legal norms, creates a space for the exercise of autonomy for subjects to manage their lives in society (Silva; Rodrigues, 2019, p. 2983). Moreover, within the spectrum of artificially intelligent beings, it will be incumbent upon them to disagree and contest the norms to create narratives of their existence vis-à-vis society and the Law. "The critical potential of the indeterminacy of legal norms, therefore, means exploring the possibility of constant revision of legal meanings" (Silva; Rodrigues, 2019, p. 2983). The function of the subject of Law, in this case, is to act as the center of the democratic legal order for the exercise of their freedoms against the practice of being ruled.

The foundation of this protection lies in the assertion that even if conceptually there is the recognition of thoughts that we do not have the exact language to describe them, denying them would be incorrect. It focuses, therefore, on an analysis beyond subjectivity in the legal construction to reconcile the private self and the objective self, from the perspective that the world without a center does not depend on the vision we have of it, nor on any other vision, but to recognize that the vision we have depends on the world put before us. Furthermore, it is this that will help the development and the balance between objectivity and subjectivity.

It must be emphasized that contemporary Law functions as a utilitarian tool to serve personal interests that do not even look at the consequences for the social, institutional framework. In our view, this is nothing more than a phantasmagoric reproduction of the Will Theory (*Willenstheorie*), which decayed in the Industrial Revolution. This theory believes that the internal will of the agent should be investigated to express his genuine intention to

protect individual interests. Its most important use was in the legal business scenario. The traditional conception of *Willenstheorie* makes the valid will of the declarant, his interests, prevail in all cases as absolute dogma to the detriment of the negotiating expectations that may have been created in the addressee or the repercussions affected by the conduct in question.

To work with the concept of autonomy without a reasonable conceptual domain is to implode the legal system itself. Not least because its elaboration presupposes knowledge of one's being, the notion of responsibility, and ethical guidelines for institutions. The lack of mastery tends to widen the possibilities of social experiences and hinder the application of the Law. This is perhaps the most significant reason for contemporary legal inefficiency. The contingency resulting from scientific and academic production is not limiting the advance of individualism and the increase of complexity since they are causes of this. Placing the individual, the biological and subjective self, as the center of the world to make the legal system revolve around him, without any ontological conceptual criteria and reflexive assumptions open doors to complexify Law and make it unenforceable since individual interests tend to conflict and what should be objective becomes solipsistic. For this reason, the construction of this topic is of importance. Tracing what role subjectivity plays in objectivity and locating the individual in the world is, above all, the basis for legal evolution.

When we insert artificial intelligence into this argument, we must analyze its evolutionary factor vis-à-vis society. The fulfillment of the emancipatory process and interspecies recognition requires a linguistic syntactic and semantic domain similar to the human level. Furthermore, it is for this reason that not all artificial intelligence technologies will be able to achieve this result quickly. In a brief classification, Bostrom (2018) assumes the existence of three stages of AI automation: 1) Artificial Narrow Intelligence (ANI); 2) Artificial General Intelligence (AGI); and 3) Artificial Superintelligence (ASI). ANI is the computational ability to efficiently perform singular tasks, such as page tracking or playing chess (Bostrom, 2020). AGI attempts to represent the original concept of intelligence by translating into algorithms that perform equivalent or superior to humans and are characterized by a deliberately programmed competence in a single narrow domain. Such modern AI algorithms tend to resemble almost all biological life (Bostrom, 2011). Moreover, finally, ASI is presented as "an intellect that far exceeds the cognitive performance of humans in virtually every domain of interest".

In the contemporary technological context, one can only detect the insertion of ANI in the informational society. The guidelines and general precepts for implementing AGI and ASI are in apparent development through machine learning and deep learning techniques (GOLDBERG; HOLLAND, 2015; CERKA et. al., 2015). It is very optimistically estimated that AGI will be available only in 2029 and that ASI would become a singular event in 2045 (Reedy, 2017). However, this does not reflect most scientists, who tend to believe that AGI will be achieved only around 2100, and ASI after 30 years of AGI discovery.

This observation leads us to two possible conclusions. The first is that the idea of recognizing autonomy to artificial intelligence at the current stage of science is instrumentalist. The goal would be to avoid the civil liability of the developer or the person responsible for it. It seems that there are no current possibilities about an AGI at a quasi-human level or with different points of view sufficient to authorize its autonomous entry into the legal order.

However, the second observation is about the possibility of recognizing artificial intelligence as a subject of Law. Affirming its impossibility at the present stage does not mean that the singularity will not be possible in the near or distant future. When this situation materializes, we will have the possibility of recognizing artificial intelligence as a subject of Law. This is so because the proceduralization of Law as a social being is not reduced to the rational factor alone. As Law is a social and cultural construction, the consequences of the insertion of a new being in the category of the subject of Law must recognize the others already framed as such. The norm is capable of disintegrating the status quo of things and emancipating the AI. Still, legal institutions alone are not capable of necessarily guaranteeing the autonomy of these entities before society.

Claims for interspecies recognition are necessary. With the term interspecies, we intend to avoid the concept of intersubjectivity made by Honneth (2017) in Struggle for Recognition[12]. To date, it is not known what the ontology of the mind is. In simpler terms, it is not known with a high

12 "In particular, Honneth's intersubjective concept of autonomy is argued to provide a normative and empirical standard for emancipation premised on the historically progressive expansion of attitudes of recognition, born out of social struggles, toward the ideal institutionalisation of mutual recognition in world politics" (Brincat, 2015, p. 225). Even in this way, Honneth's intersubjectivity is tied to the human condition. In our work, we want to untie this condition and grant to subject the condition of being recognized in the law.

degree of certainty what the reason is. It is assumed that subjectivity is amalgamated with the biological aspect and that only biological beings with a brain can produce subjectivity (Searle, 1980). Therefore, the intersubjective terminology recognition seems to be erroneous to apply in this scenario.

In this sense, the view from nowhere developed by Nagel is essential. It must be recognized that the world exists independently of an anthropomorphic and subjective conception. Moreover, different views can exist that enhance an objective reality without necessarily being subjective. Furthermore, this is the case with Artificial Intelligence. Moreover, as a final argument, it is verified that Law does not require subjectivity for the constitution of the subject of Law. This statement becomes evident from the formation of corporations as fictitious subjects in the legal sphere from a cultural demand and requirement destined to the patrimonial protection of the subjects who exercise the entrepreneurial activity.

The second point is that Law is a product of social forces, human activity (Neumann, 2013, p. 72). The legitimacy to emancipate a being can be achieved in institutional foster care from objective claims without destroying its form of legal imputation. Subjectivity is not rejected. Not least because any worldview in its most objective aspect must recognize the first-person speeches for its structure to be complete. Therefore, any objective view that rejects subjectivity seems to be in error because subjectivity is part of the world (Nagel, 1986). Nevertheless, in reality, every subjective view that rejects the possibility of other objective forms of participation is also mistaken[13].

Thus, at the core of the rationality of Law is the person as the center of imputation for domination and for the possibility of democratic participation in the norms that govern his life (Silva; Rodrigues, 2019, p. 2983). On a more abstract level, interspecies recognition is a complementary step to the emancipatory process that guarantees the non-human entity (the AI, in this case) freedom from the state and society to perform acts that require responsibility and to be able to form themselves as political beings. For an

13 The procedural theory of the subject of law is not only the opposite version of subjectivity and objectivity but a merge between them. In a conceptual framework, the gains by using this theory are detected by the use of rationality, which cannot be defined as objective and subjective as well. If you ask a child how much is 2 plus 2 and she answers 5, you cannot say that she does not have any subjectivity, but their behavior is irrational. Even human beings (sentients) can be irrational many times. The ideal criteria used in our theory if focused on rationality, which represents sometimes the qualia and the objective part of the world. After all, we can detect the condition of the subject of law for those who want and is capable of fighting against the system.

AI to be considered a subject of Law, society must recognize and support for this to occur so that the emancipatory process gains strength from the realization of linguistic mastery and the practice of not being governed. This implies the need for an AI to act in the same way (or better) as a human being, a political being.

4 FINAL CONSIDERATIONS

Given the above, the following considerations can be made:

1. It is unintelligible to understand "What is it like to be artificial intelligence?"
2. Different world views can compose objective reality.
3. As a subjective science, Law needs to recognize other points of view besides those of human beings, under penalty of falling into the falsehood of reality.
4. The formation of the subject of Law necessarily goes through the emancipatory process, which artificial intelligence can face, but not now.
5. This means that it is possible to consider artificial intelligence as a subject of Law, as long as it is AGI or ASI.
6. The recognition of rights and duties outside this scenario is nothing more than instrumentalism based on physicalism.

Therefore, it must be recognized that the world exists independently of an anthropomorphic and subjective conception. Moreover, there can be different views that enhance an objective reality without necessarily being subjective. Furthermore, this is the case with Artificial Intelligence. It is verified that Law does not require subjectivity for the constitution of the subject of Law. This statement becomes evident from the formation of corporations as fictitious subjects in the legal sphere from a cultural demand and requirement to protect the assets of the subjects who exercise business activities.

Given the above, we conclude that it is unintelligible to understand what it is like to be artificial intelligence. Therefore, it is also unintelligible to attribute an autonomous regulation to these entities in the current scenario due to the lack of scientific and technological progress. However, this does not mean that these entities can be left aside and their worldviews denied

by the possible absence of subjectivity. With the procedural theory of the subject of Law, artificial intelligence can be elevated to a social and legal level similar to the human one, as long as it demonstrates this understanding to deserve its legal protection.

REFERENCES

BERKLEY, George. *A treatise concerning the principles of Human Knowledge*. 1710, seções 2 e 23. Available in: https://www.earlymoderntexts.com/assets/pdfs/berkeley1710.pdf. Accessed 10 aug. 2021.

BOSTROM, N. *Superinteligência*. Rio de Janeiro: Darkside, 2018.

_____. The ethics of artificial intelligence. In RAMSEY, W.; FRANKISH, K. (Org.) *Draft for Cambridge Handbook of Artificial Intelligence*. Cambridge University Press, 2011.

BRINCAT, Shannon. The harm principle and recognition theory: On the complementarity between Linklater, Honneth and the project of emancipation. *Critical horizons*, v. 14, n. 2, p. 225-256, 2013.

ČERKÃ, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liãbility for dãmãges cãused by Artificiãl Intelligence. *Computer Law & Security Review*, Elsevier, v. 31, n. 3, p. 376-389, jun. 2015.

DIVINO, S. B. S. Inteligência Artificial como sujeito de direito: construção e teorização crítica sobre pessoalidade e subjetivação. *Revista de Bioética y Derecho*, v. 52, p. 237-252, 2021.

_____. Procedural theory of the subject of law and non-human animals: criteria for recognition of legal subjectivity from the perspective of critical theory. *Revista Brasileira de Políticas Públicas,* Brasília, v. 10, n. 3. p. 181-195, 2020.

FOUCAULT, M. *Em defesa da sociedade*. Curso no Collège de France (1975-1976). São Paulo: Martins Fontes, 1999.

_____. *O que é a crítica?* Seguido de A cultura de si. Lisboa: Texto & Grafia, 2015.

_____. O sujeito e o poder. In: RABINOW, P.; DREYFUS, H. L. (Org.). *Michel Foucault*: uma trajetória filosófica (para além do estruturalismo e da hermenêutica). Rio de Janeiro: Forense Universitária, 1995. p. 231-249.

GOLDBERG, D. E.; HOLLAND, J. H. Genetic algorithms and machine learning. In *Machine learning*. Switzerland: v. 3, 1988, p. 95-99.

HERVADA, J. *O que é o direito*? A moderna resposta do realismo jurídico. São Paulo: Martins Fontes, 2006.

HONNETH, A. *Luta por reconhecimento*. São Paulo: Editora 34, 2017.

KELSEN, H. *Pure theory of law*. London: Cambridge University Press, 1967.

LUHMANN, N. Sociologia do Direito. Rio de Janeiro: Tempo Brasileiro, v. I, 1983.

MARTINS-COSTA, J. *Indivíduo, pessoa, sujeito de direitos:* contribuições renascentistas para uma história dos conceitos jurídicos. Philia&Filia. 1, 1, 69-95, 2010.

MCCARTHY, J. *What is artificial intelligence?* Stanford University, 2007.

NAGEL, T. *The view from nowhere.* New York: Oxoford University Press, 1986.

_____. What is it like to be a Bat? In: *The Philosophical Review.* Duke University Press, v. 83, n. 4 (Oct.), p. 438-439, 1974.

NEUMANN, Franz. O conceito de liberdade política. *Cadernos de Filosofia Alemã*, São Paulo, n. 22, p. 107-154, 2013.

REEDY, C. Kurzweil Claims That the Singularity Will Happen by 2045. Futurism. 2017. Available in: https://futurism.com/kurzweil-claims-that-the-singularity-will-happen-by-2045. Acessed 10 aug. 2021

RUSSELL, Stuart. J.; NORVIG, Peter. *Artificial intelligence*: a modern approach. 3. ed. New Jersey: Pearson Education, 2010.

SEARLE, J. Minds, brains, and programs. *Behavioral and Brain Sciences*, v. 3, n. 3, p. 417-424, 1980.

SILVA, S. S.; RODRIGUEZ, J. R. Para que serve ser uma pessoa no Direito? Diálogos no campo crítico. *Rev. Direito Práx.*, Rio de Janeiro, v. 10, n. 4, p. 2968-3023, 2019.

VIOLA, F. O estatuto jurídico da pessoa em perspectiva histórica. *Revista da Faculdade de Direito da UFRGS,* Porto Alegre, n. 36, v. esp., p. 12-29, out. 2017.

Sobre os autores:

Sthéfano Bruno Santos Divino | *E-mail:* sthefanoadv@hotmail.com

Doutorando (2020 – Bolsista do Programa de Excelência Acadêmica – Proex – Capes/Taxa) e Mestre (2019) em Direito Privado pela Pontifícia Universidade Católica de Minas Gerais. Bacharel em Direito pelo Centro Universitário de Lavras (2017). Professor Adjunto do Curso de Direito do Centro Universitário de Lavras (2020 – atual). Professor substituto de Direito Privado da Universidade Federal de Lavras (03/2019 – 03/2021). Advogado.

Rodrigo Almeida Magalhães | *E-mail:* amagalhaes@ig.com.br

Doutor e Mestre em Direito pela Pontifícia Universidade Católica de Minas Gerais. Professor da Graduação na PUC Minas e na Universidade Federal de Minas Gerais. Professor do Mestrado e Doutorado da PUC Minas.

Data de submissão: 18 de agosto de 2021.
Data de aceite: 10 de janeiro de 2022.

Argumentação Jurídica e Aprendizado Profundo

Legal Argumentation and Deep Learning

ORLANDO LUIZ ZANON[1]
Universidade do Vale do Itajaí (Univali).

GUILHERME KIRTSCHIG[2]
Universidade do Vale do Itajaí (Univali).

RESUMO: O objetivo geral deste artigo consiste em avaliar a possibilidade de automatização da argumentação jurídica, mediante uso da tecnologia de aprendizado profundo. Como objetivos específicos, busca-se, primeiro, apresentar um panorama da argumentação jurídica e sua importância na aplicação do Direito, a partir das concepções vinculadas a distintos paradigmas da ciência jurídica; segundo, discorrer sobre o aprendizado profundo, considerando sua concepção, características e, também, aplicabilidade no campo jurídico; bem como, terceiro, problematizar o efetivo desempenho da argumentação jurídica por parte dos robôs, sopesando suas exigências com as limitações inerentes ao aprendizado profundo. Em considerações finais, apresenta-se argumento quanto à possibilidade de compatibilização entre ambos, mediante a inserção dos dados emergentes do aprendizado profundo na atividade de aplicação do Direito. A pesquisa encetada operou com o método indutivo em sua fase de investigação, o método analítico na fase de tratamento de dados e, no presente relatório em forma de artigo científico, utiliza-se novamente o método indutivo.

PALAVRAS-CHAVE: Aplicação do Direito; argumentação jurídica; inteligência artificial; aprendizado profundo; decisão judicial.

ABSTRACT: This article aims to evaluate the possibility of automating legal argumentative practice using deep learning technology. As specific objectives, it seeks to present an overview of the legal argumentative practice and its importance for the application of the Law, through approaches linked to different legal paradigms; to discuss deep learning, its concept and characteristics, and its incursion in the application of the Law; problematize the effective performance of legal argumentation by robots, balancing its requirements with the limitations inherent to deep learning. In final considerations, the insertion, in the legal discourse, of data emerging from deep learning, is presented as a possibility of reconciling both. The research started with the inductive method in its investigation phase, used

1 Orcid: https://orcid.org/0000-0002-0021-9278.
2 Orcid:

the analytical method in the data treatment phase, and, in the present report, the inductive method was used again.

KEYWORDS: Law adjudication; legal argumentation; artificial intelligence; deep learning; judicial decision.

SUMÁRIO: Introdução; 1 A argumentação e sua função na aplicação do Direito; 2 O aprendizado futuro e sua incursão no Direito; 3 Podem os robôs argumentar juridicamente?; Considerações finais; Referências.

INTRODUÇÃO

O presente artigo se propõe a discutir uma questão relacionada à automatização da aplicação do Direito mediante a utilização da tecnologia da inteligência artificial, mais especificamente aquela denominada aprendizado profundo (*deep learning*), considerando as peculiaridades da argumentação jurídica.

Busca-se esclarecer se a produção de arrazoados jurídicos, por robôs dotados dessa tecnologia, efetivamente consiste no desempenho da atividade argumentativa inerente à aplicação do Direito, bem como visa se argumentar sobre as formas de compatibilizar suas limitações com as funções e exigências da argumentação jurídica.

Para tanto, inicialmente são tecidas considerações sobre a argumentação jurídica e sua proeminência, a partir das concepções de doutrinadores ligados a distintos paradigmas da ciência jurídica.

A seguir, discorre-se sobre o aprendizado profundo, considerando sua concepção, características e, também, aplicabilidade no campo jurídico, tendo em vista a atividade argumentativa.

Por fim, problematiza-se o efetivo desempenho dessa atividade por parte dos robôs e, sopesando-se as exigências da argumentação jurídica com as limitações inerentes à aludida tecnologia, argumenta-se quanto a um modo de compatibilizá-las.

A pesquisa encetada operou com o método indutivo em sua fase de investigação, o método analítico na fase de tratamento de dados e, no presente relatório, em forma de artigo científico, utiliza-se novamente o mé-

todo indutivo. As técnicas de suporte foram a da categoria, do conceito operacional, do referente e da pesquisa bibliográfica[3].

1 A ARGUMENTAÇÃO E SUA FUNÇÃO NA APLICAÇÃO DO DIREITO

Conquanto possam existir divergências relativamente à função da argumentação no âmbito do Direito, a depender do paradigma[4] no qual se insira determinada visão do fenômeno jurídico, não parece haver dissenso quanto à importância dessa atividade para a aplicação jurídica.

Chaim Perelman e Lucie Olbrechts-Tyteca[5] afirmam que argumentar significa "influenciar, por meio do discurso, a intensidade de adesão de um auditório a certas teses", algo típico na atividade jurígena.

De acordo com Luigi Ferrajoli, proponente de uma visão positivista do Direito[6], na argumentação reside a solução geral e abstrata para os conflitos e incertezas que surgem na aplicação e na interpretação das normas jurídicas, pois, com ela, é possível sustentar "a qualificação jurídica proposta como a mais correta com base nas normas existentes"[7]. Ele acrescenta ainda que "toda atividade jurisprudencial, exatamente pelo fato de ser submetida à lei e, portanto, ao ônus da motivação, consiste em uma atividade argumentativa, além de aplicativa"[8].

A argumentação permite aos operadores jurídicos motivar, externar e comunicar suas interpretações das prescrições validamente produzidas pelas autoridades competentes, atribuindo-lhes significados normativos[9]. Isso

3 Sobre métodos e técnicas, vide PASOLD, Cesar Luiz. *Metodologia da pesquisa jurídica* – Teoria e prática. 14. ed. rev., atual. e ampl. Florianópolis: EMais, 2018. p. 89 a 115.

4 Orlando Luiz Zanon Júnior, a partir das ideias de Thomas Kuhn, explica que as proposições científicas, destinadas a resolver diversos problemas, têm apoio em bases teóricas compartilhadas, que proporcionam conceitos, métodos e instrumentos. Enquanto essas bases forem adequadas para tal objetivo, elas se mantêm, constituindo um paradigma. O surgimento de uma questão insolúvel gera uma crise no paradigma, propiciando o surgimento de novos modelos teóricos que proporcionem a solução desses novos problemas. A partir da consolidação dos novos modelos teóricos, constitui-se um novo paradigma, que perdurará até que seja encontrada nova anomalia. Vide ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. 3. ed. rev. e ampl. São Paulo: Tirant lo Blanch, 2019. p. 36 a 42.

5 PERELMAN, Chaim; OLBRECHTS-TYTECA, Lucie. *Tratado da argumentação*. A nova retórica. Trad. Maria Ermantina de Almeida Prado Galvão. São Paulo: Martins Fontes, 2005. p. 16. Título original: *Traité de l'argumentation*.

6 Sobre essa caracterização, vide ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 67.

7 FERRAJOLI, Luigi. *A democracia através dos direitos*. O constitucionalismo garantista como modelo teórico e como projeto político. Trad. Alexander Araújo de Souza e outros. São Paulo: Revista dos Tribunais, 2015. p. 135.

8 FERRAJOLI, Luigi. Op. cit., p. 137.

9 FERRAJOLI, Luigi. Op. cit., p. 138 e 139.

resulta no que Ferrajoli[10] denomina direito vivente, uma dimensão performativa do fenômeno jurídico, que compõe sua essência juntamente com o direito vigente, ao qual está jungida toda atividade discursiva.

Em polo antagônico à visão juspositivista, Ronald Dworkin[11] aponta que a verdade de uma proposição valorativa, dentre as quais ele inclui aquelas atinentes à aplicação do Direito, depende da construção de uma argumentação adequada, assim entendida aquela densa, complexa, íntegra e apoiada em outras esferas do universo valorativo, com as quais devem guardar harmonia e coerência. A construção de uma argumentação nesses moldes é, para ele, uma questão de integridade e responsabilidade moral (e, mais precisamente, política)[12].

Para além dessa harmonia em relação a outros domínios do universo valorativo, Dworkin vislumbra também a necessidade de uma coerência interna ao próprio âmbito do Direito. A argumentação construída para a solução de um caso concreto deve permitir tanto a justificação vertical de sua decisão, ou seja, a coerência em relação às decisões adotadas por órgãos jurisdicionais superiores, propiciando seu controle, como também a horizontal, enquanto consistência com decisões do mesmo nível[13].

Como explica Vera Karam de Chueiri[14], ao tratar do pensamento do referido autor, ele exemplifica sua visão da aplicação jurídica como "um romance em cadeia, no qual vários autores escrevem um romance em conjunto. O romance deve ficar o mais coerente possível".

Não se pode olvidar, ainda, que Dworkin sustenta existir uma única resposta certa para cada caso (*the one right answer*), a qual não é dada, mas construída argumentativamente, de forma a inserir-se coerentemente no romance em cadeia, do qual cada decisor é um dos autores[15].

10 FERRAJOLI, Luigi. Op. cit., p. 138 e 139.

11 DWORKIN, Ronald. *Justice for hedgehogs*. Cambridge: Harvard University Press, 2011. p. 99 a 122.

12 DWORKIN, Ronald. *Justice for hedgehogs*. Op. cit., p. 99 a 122.

13 DWORKIN, Ronald. *Levando os direitos a sério*. Trad. Nelson Boeira. São Paulo: Martins Fontes, 2002. p. 182 e 183. Título original: *Taking rights seriously*.

14 KARAM DE CHUEIRI, Vera. *A filosofia jurídica de Ronald Dworkin como possibilidade de um discurso instituinte de direitos*. Dissertação (Mestrado em Direito) – Universidade Federal de Santa Catarina. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/106357/90511.pdf?sequence=1&isAllowed=y. Acesso em: 1º fev. 2021, p. 56 e 57.

15 KARAM DE CHUEIRI, Vera. Op. cit., p. 64.

Como se percebe, Dworkin, na qualidade de pós-positivista da linha substancialista[16], vislumbra um papel criativo na interpretação do Direito, ainda que limitado pela necessidade de manutenção da integridade da construção coletiva[17]. Ao interpretar, o aplicador está criando mais um capítulo da história jurídica em desenvolvimento, mantendo conexões com o passado e estabelecendo bases para as construções futuras[18].

Também para autores da linha pós-positivista de viés procedimentalista, como Manuel Atienza, a prática do Direito consiste, "fundamentalmente, em argumentar"[19].

Robert Alexy, considerado um dos principais autores dessa linha de pensamento[20], destaca o caráter legitimador da argumentação jurídica, tanto no tocante à validade das normas jurídicas quanto no papel do Direito como instrumento de resolução de conflitos sociais.

Em relação à primeira (argumentação), o autor[21] assenta sua dependência em relação a uma pretensão de correção, por sua vez aferível a partir da fundamentação racional e intersubjetiva das proposições de aplicação das normas, bem como da adequação dessa motivação aos consensos atingidos na experiência jurídica de determinada comunidade.

Quanto à resolução de conflitos sociais, Alexy[22] afirma que as decisões judiciais contemplam os denominados julgamentos de valor, ou seja, afirmações no sentido de que um curso de ação é preferível em relação a outro.

Cabe à argumentação justificar racionalmente esses posicionamentos[23], fechando a brecha entre a lei escrita e uma justa solução dos problemas legais, "de acordo com os padrões da razão prática e dos conceitos de justiça bem fundamentados da comunidade"[24].

16 Vide ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 90.

17 KARAM DE CHUEIRI, Vera. Op. cit., p. 94 e 95.

18 KARAM DE CHUEIRI, Vera. Op. cit., p. 94 e 95.

19 ATIENZA, Manuel. As razões do Direito. *Teorias da argumentação jurídica*. Trad. Maria Cristina Guimarães Cupertino. São Paulo: Landy, 2003. p. 17. Título original: *Las razones del derecho. Teorías de la argumentación jurídica*.

20 Vide ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 90.

21 ALEXY, Robert. *Conceito e validade do Direito*. Trad. Gercélia Batista de Oliveira Mendes. São Paulo: Martins Fontes, 2011. p. 58-67. Título original: *Begriff und Geltung des Rechtes*.

22 ALEXY, Robert. *Teoria da argumentação jurídica*. Trad. Zilda Hutchinson Schild Silva. São Paulo: Landy, 2001. p. 20 e 21. Título original: *Theorie der juristische argumentation*.

23 ALEXY, Robert. *Teoria da argumentação jurídica*. Op. cit., p. 20 e 21.

24 ALEXY, Robert. *Teoria da argumentação jurídica*. Op. cit., p. 34.

Nessa linha lógica, Alexy propõe-se à construção de uma teoria da argumentação jurídica, com vistas a proporcionar um critério "para julgar a correção de afirmações normativas, como instrumento crítico para excluir tudo o que não seja racional numa justificação objetiva, e/ou por tornar mais visível um ideal pelo qual valha a pena lutar"[25]. Portanto, o argumento racional, construído de acordo com a teoria proposta, permite aquilatar a correção de uma proposição de aplicação do Direito[26].

Na sequência, cabe trazer à baila o papel da argumentação no pós-positivismo de corte pragmático, segundo as considerações de Richard Posner, seu autor mais conhecido.

Segundo Posner[27], o pragmatismo jurídico consiste em uma disposição para fundar as escolhas empiricamente, em fatos e consequências, em lugar de abstratamente, em conceitos gerais. Na sua visão[28], o pragmatismo pressupõe a inviabilidade de justificações lastradas em princípios morais elevados, que possam servir como guia absoluto para decisões nos campos político e jurídico. Outrossim, as decisões judiciais merecem ser criticadas apenas se baseadas em fatos erroneamente apreendidos ou em consequências ruins.

Todavia, como o próprio Posner[29] esclarece, nas escolhas realizadas pelos aplicadores do Direito, devem ser consideradas tanto as prováveis consequências sistêmicas que delas advirão, como também os resultados concretos, específicos da situação sob exame. Naquelas (consequência sistêmicas) se inclui a preservação dos valores contidos nas normas jurídicas, que se expressam nos princípios emanados dos textos constitucionais, legais e na jurisprudência, pois há inegáveis benefícios em restringir o arbítrio, aumentar a transparência e simplificar a aplicação do Direito[30].

Ocorre que a indeterminação, a ambiguidade, a indefinição e a ausência de propósitos orientadores não são excepcionais na aplicação do

25 ALEXY, Robert. *Teoria da argumentação jurídica*. Op. cit., p. 29.

26 ALEXY, Robert. *Teoria da argumentação jurídica*. Op. cit., p. 273.

27 POSNER, Richard. *Direito, pragmatismo e democracia*. Trad. Teresa Dias Carneiro. Rio de Janeiro: Forense, 2010. p. 1 a 10 e 38 a 44. Título original: *Law, pragmatism and democracy*.

28 POSNER, Richard. Op. cit., p. 1 a 10 e 38 a 44.

29 POSNER, Richard. Op. cit., p. 49 a 58 e 60 e 61.

30 POSNER, Richard. Op. cit., p. 49 a 58.

Direito[31], de modo que somente a avaliação de consequências sistêmicas não será suficiente para a tomada de uma decisão.

Por essa razão, toda questão importante pode ser usada para embasar uma decisão jurídica acerca de um determinado problema, sendo que justamente essa abertura abre espaço para a criatividade, *pari passu* com a continuidade[32]. Nesse sentido, a aplicação pragmática do Direito ostenta orientação empiricista interdisciplinar, ou seja, aberta "a invasões ao direito vinda de outros domínios do saber"[33].

O autor aponta que incumbe ao aplicador buscar o equilíbrio entre esses diversos elementos empíricos e sistêmicos, cabendo à argumentação externar esse sopesamento de prós e contras, em busca da decisão mais razoável possível, no tocante às consequências que engendrará[34].

De outra margem, cabe referir a Teoria Complexa do Direito, proposta por Orlando Luiz Zanon Junior, consistente em uma proposição mais do que positivista de paradigma para a ciência jurídica, no tocante às suas quatro plataformas elementares, consistentes nas teorias das fontes, do ordenamento, da norma e da decisão[35].

Zanon[36] enxerga na argumentação o próprio material com o qual são construídas as normas jurídicas, a partir de um problema fático, mediante um empreendimento decisório lastrado em fontes jurídicas (elementos de determinação) e provas (elementos de aproximação fática).

Segundo o autor[37], o sistema jurídico não contém normas abstratas completamente prontas para mera incidência subsuntiva, como se pairando no éter à espera de aplicação em casos futuros. Diversamente, há elementos concretamente aferíveis por observação (as fontes, como textos legais, jurisprudência etc.) que servem como guia para que o aplicador, diante do complexo probatório, construa a resposta adequada à resolução do caso (a norma jurídica), mediante esforço argumentativo de justificação.

31 POSNER, Richard. Op. cit., p. 53.
32 POSNER, Richard. Op. cit., p. 10 e 50.
33 POSNER, Richard. Op. cit., p. 58.
34 POSNER, Richard. Op. cit., p. 50.
35 ZANON JUNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit.
36 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 184 a 189.
37 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 183 a 201.

A argumentação é a forma de articulação dos elementos decisórios e fáticos, por meio de um empreendimento voltado à resolução do conflito, caracterizando a atividade de produção do Direito[38].

Zanon[39] assenta que a norma jurídica é sempre uma resposta única, resultado de uma operação cognitiva complexa, de modo que novas situações demandarão atividade de interpretação específica, para fins de articular os elementos de determinação com as provas, sem que esteja afastada a invocação de parâmetros interdisciplinares.

Essa tarefa constitui uma "ponte entre o ser e o dever-ser"[40], caracterizada por uma pretensão de verdade transitória, argumentativa, passível de demonstração intersubjetiva e discursiva[41].

Ele acrescenta que a Teoria Complexa do Direito adota uma versão fraca da teoria da "única resposta correta", pois essa correção está calcada na mencionada verdade transitória[42].

Após essa síntese de abordagens diferenciadas quanto às compreensões do Direito e à atuação dos seus intérpretes e aplicadores, examinadas acima, verifica-se a existência de uma grande carga depositada sobre a atividade argumentativa, porquanto responsável por concretizar a aplicação jurídica nos quadros de cada uma dessas teorias.

Justamente no âmbito dessa atividade é possível observar o ingresso de inovações tecnológicas que afetam o seu desempenho, algumas das quais com efeitos questionáveis em relação às suas variadas funções. O item seguinte examinará uma dessas inovações: o aprendizado profundo.

2 O APRENDIZADO FUTURO E SUA INCURSÃO NO DIREITO

A reconfiguração das relações sociais em torno das tecnologias da informação, do processamento de dados e da comunicação, a partir dos anos finais do século XX, impactou todos os campos da atividade humana, consistindo em autêntica revolução[43].

38 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 183 a 201.
39 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 183 a 201.
40 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 189.
41 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 195 e 196.
42 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 195 e 196.
43 CASTELLS, Manuel. *A sociedade em rede*. Trad. Roneide Venâncio Majer. São Paulo: Paz e Terra, v. I, 2002. p. 67 e 68. Título original: *The rise of the network society*.

Trata-se de movimento abrangente, pois abarca diversas áreas do conhecimento e ambiciona atuar sobre os grandes desafios da sociedade, por meio de soluções multidisciplinares[44].

Certamente que o Direito não poderia deixar de ser colhido por essa "força das coisas"[45]. A aplicabilidade das novas tecnologias informacionais ao direito vem ganhando impulso e suscitando crescente atenção e reflexão, especialmente no tocante à inteligência artificial (IA), dotada de grande potencial disruptivo nessa área[46].

Segundo Fabiano Hartmann Peixoto e Roberta Zumblick Martins da Silva, a IA consiste em uma "subárea da ciência da computação e busca fazer simulações de processos específicos da inteligência humana por intermédio de recursos computacionais"[47].

Por sua vez, também a IA se divide em subáreas, uma das quais é o aprendizado de máquina (*machine learning*), voltado à habilitação de computadores para aprenderem sozinhos a executar tarefas, para as quais não sejam especificamente programados[48].

O aprendizado profundo (*deep learning*) é uma modalidade de aprendizado de máquina, na qual os computadores aprendem tarefas complicadas a partir de sua própria experiência com outras mais simples, por intermédio de redes de sistemas informacionais interconectadas, chamadas de redes neurais, em virtude de serem dispostas como neurônios biológicos[49].

O desenho e o funcionamento das redes neurais artificiais são inspirados em recentes descobertas acerca do *modus operandi* do cérebro humano, as quais apontam que as funções dos sistemas sensoriais, motores e cognitivos são executadas por inúmeros neurônios atuando em paralelo, formando redes estratificadas em distintas profundidades[50].

---

44 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, v. I, 2019. p. 50.

45 Expressão utilizada por Jânia Maria Lopes Saldanha, Rafaela da Cruz Mello e Têmis Limberger para denominar o fenômeno através do qual uma representação dominante do mundo, em determinada época, altera a realidade política, social e econômica, e acaba por atingir o Direito. Vide SALDANHA, Jânia; MELLO, Rafaela da Cruz; LIMBERGER, Têmis. Do governo por leis à governança por números: breve análise do *trade in service agreement* (*TISA*). *Revista de Direito Internacional*, v. 13, n. 3, p. 338 a 355, 2016, p. 346.

46 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 58.

47 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 20 e 21.

48 SEJNOWSKI, Terrence J. *A revolução do aprendizado profundo*. Trad. Carolina Gaio. Rio de Janeiro: Alta Books, 2019. p. 299. Título original: *The deep learning revolution*.

49 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 97.

50 SEJNOWSKI, Terrence J. Op. cit., p. 41 a 44.

No cérebro humano, as informações de entrada, captadas pelos sentidos, são desmembradas e processadas por diversas camadas de neurônios em rede, conforme padrões característicos que apresentem e sejam estatisticamente relevantes em seu ambiente[51]. Conforme a importância desses padrões, sua detecção em uma camada neuronal ativará a camada seguinte, cuja entrada será a saída da camada anterior, e assim por diante, até o resultado final de saída, que embasa as tomadas de decisão humanas[52].

O aprendizado ocorre por meio da formação de sinapses, ou seja, áreas de interação entre os neurônios, moduladas conforme as necessidades de solução dos problemas que se apresentam, a partir do reconhecimento de padrões e estabelecimento de associações[53].

Terrence Sejnowski pontua que o propósito é essencial para o desenvolvimento dessas conexões, não havendo aprendizado sem um problema a solucionar, bem como acrescenta que, embora os neurônios de cada ser humano permaneçam basicamente os mesmos durante a vida, suas sinapses mudam constantemente, estando essa plasticidade associada às necessidades que continuamente se renovam e modificam[54].

As redes neurais artificiais emulam esse mecanismo, por meio de algoritmos que atribuem pesos a determinados padrões e associações detectados nas unidades de entrada[55]. Os pesos são "medidas de influência que cada entrada tem na decisão final feita pela unidade de saída"[56]. Uma vez que uma camada da rede atinja determinado patamar de ativação, conforme esses pesos, o algoritmo aciona a camada seguinte, formando conexões entre as unidades da rede, ou seja, sinapses artificiais[57].

O aprendizado ocorre provendo-se uma base de exemplos à rede e criando-se algoritmos que ajustem os pesos conforme os resultados da saída

51 SEJNOWSKI, Terrence J. Op. cit., p. 41 a 49.

52 SEJNOWSKI, Terrence J. Op. cit., p. 41 a 49.

53 Conceito de sinapse conforme o *Dicionário Michaelis online*. Sinapse. Disponível em: https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=sinapse. Acesso em: 1º fev. 2021. Sobre a importância das sinapses para o aprendizado, ver SEJNOWSKI, Terrence J. Op. cit., p. 65 e 66 e 74 a 77.

54 SEJNOWSKI, Terrence J. Op. cit., p. 68.

55 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 98. O algoritmo é o "[...] *esquema ejecutivo de la máquina, almacenando todas las opciones de decisión em función de los datos que se vayan conociendo*". Cf. NIEVA FENOLL, Jordi. *Inteligencia artificial y processo judicial*. Madrid: Marcial Pons, 2018. p. 21.

56 SEJNOWSKI, Terrence J. Op. cit., p. 44.

57 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 98 e 99.

estejam corretos ou não[58]. Desse modo, executa-se o treinamento da máquina, que extrapola para atuação em casos mais elaborados do que aqueles inicialmente fornecidos.

Para além dessa concepção básica, outras técnicas foram adicionadas para compor os modelos atuais de redes neurais, com vistas à consecução de seus objetivos de solucionar problemas ao modo da inteligência humana[59].

Como já indicado, o aprendizado profundo não é a única solução de IA existente, nem a única modalidade de aprendizado de máquina possível. No entanto, Sejnowski considera que o seu desenvolvimento é revolucionário, por romper com uma visão de IA baseada em processamento sequencial de símbolos por meio de regras lógicas, que tornava os sistemas computacionais excessivamente complicados, custosos e pouco eficientes[60]. O aprendizado profundo, ao contrário, proporciona um modelo de inteligência adequado à solução rápida de problemas inerentes às atividades humanas cotidianas[61].

O uso da linguagem natural é uma dessas atividades, e numerosos problemas humanos são resolvidos através de seu uso. O aprendizado profundo, juntamente com o desenvolvimento da neurociência, relaciona-se intimamente a uma outra mudança de paradigma, desta vez na linguística, em cujo âmbito a ênfase no processamento de símbolos para aquisição da linguagem foi substituída pelo aprendizado associativo, a partir da experiência e em um rico contexto de relações entre os vários elementos proporcionados pelos sentidos[62].

O modelo é atraente para o Direito, um produto da dialética entre intenção sistemática e experiência problemática[63], essencialmente vinculado ao contexto social, econômico e cultural no qual se insere, sendo expresso

58 SEJNOWSKI, Terrence J. Op. cit., p. 45 a 53.

59 Por exemplo, a retropropagação de erros, o aprendizado convolucional, o aprendizado por recompensa, ou os modos supervisionado e não supervisionado de aprendizado. Vide SEJNOWSKI, Terrence J. Op. cit., p. 87 a 184; PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 91 a 105.

60 SEJNOWSKI, Terrence J. Op. cit., p. 31 a 86.

61 SEJNOWSKI, Terrence J. Op. cit., p. 31 a 86; PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 97.

62 SEJNOWSKI, Terrence J. Op. cit., p. 266 a 272. Peixoto e Silva mencionam três ondas da relação entre IA e linguística, sendo a utilização das técnicas de aprendizado profundo a mais recente. Vide PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 82 a 84.

63 AMARAL, Francisco. Racionalidade e sistema no direito civil brasileiro. *Revista de Informação Legislativa*, Brasília, a. 31, n. 121, p. 233 a 243, jan./mar. 1994, p. 237.

por meio da linguagem. O aprendizado profundo tem sido utilizado, assim, para a construção de argumentos jurídicos a partir de uma base de dados[64].

O robô Ross (solução de IA elaborada a partir da aplicação Watson da IBM), por exemplo, opera segundo essa tecnologia e, assim, é capaz de construir argumentos em linguagem natural para um determinado problema jurídico que se lhe apresente[65].

Interessante também é o projeto de IA denominado Victor, desenvolvido em parceria entre a Universidade de Brasília (UnB) e o Supremo Tribunal Federal (STF), cujo objetivo inicial é automatização da análise textual de peças processuais, para fins de aquilatar o cumprimento dos requisitos de repercussão geral dos recursos extraordinários que são encaminhados ao tribunal[66]. Esse robô também é um exemplar de tecnologia de aprendizado profundo, estruturando-se em redes neurais, por meio das quais ele aprende a linguagem natural utilizada nos textos jurídicos sujeitos a exame de repercussão geral[67].

Embora o STF enfatize que o Victor não irá tomar decisões, mas apenas auxiliar o trabalho jurisdicional da Corte[68], registra-se também que ele irá devolver automaticamente aos tribunais de origem os processos enquadrados em algum dos temas de repercussão geral[69]. Trata-se de decidir o destino de recursos extraordinários, atividade jurisdicional típica que, embora possa estar fundamentada no enquadramento do recurso em tema de repercussão geral, na realidade consistirá em referendar uma constatação nesse sentido, alcançada pelo robô.

Não se afigura distante ou impensável, diante desse desenvolvimento, que a sociedade brasileira se defronte com um Magistrado robô dotado

64 NIEVA FENOLL, Jordi. Op.cit., p. 29 a 31 e 115.

65 NIEVA FENOLL, Jordi. Op.cit., p. 29 a 31 e 115.

66 MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Projeto Victor: perspectivas de aplicação da inteligência artificial ao Direito. *Revista Direito e Garantias Fundamentais*, v. 19, n. 3, p. 219 a 238, set./dez. 2018, p. 226.

67 MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Op. cit., p. 226; ANDRADE, Mariana Dionísio de; PINTO, Eduardo Régis Girão de Castro; LIMA, Isabel Braga de; GALVÃO, Alex Renan de Souza. Inteligência artificial para o rastreamento de ações com repercussão geral: o Projeto Victor e a realização do princípio da razoável duração do processo. *Revista Eletrônica de Direito Processual – REDP*, a. 14, v. 21, n. 1, p. 312 a 335, jan./abr. 2020, p. 323.

68 BRASIL. Supremo Tribunal Federal. *Inteligência artificial*: trabalho judicial de 40 minutos pode ser feito em 5 segundos (2018). Disponível em: http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=393522. Acesso em: 1º fev. 2021.

69 TEIXEIRA, Mateus. *STF investe em inteligência artificial para dar celeridade a processos* (2018). Disponível em: https://www.jota.info/coberturas-especiais/inova-e-acao/stf-aposta-inteligencia-artificial-celeridade-processos-11122018. Acesso em: 1º fev. 2021.

de atribuições de maior envergadura, como se tem observado em outros países[70].

Todavia, certas características do aprendizado profundo, associadas às funções exercidas pela argumentação jurídica, como anotadas no item 1, podem apresentar algumas dificuldades para o atingimento desse desiderato. Elas, junto com possíveis soluções, serão apresentadas no item subsequente.

3 PODEM OS ROBÔS ARGUMENTAR JURIDICAMENTE?

Conforme já foi possível observar no item 1, é pesado o fardo da argumentação no âmbito do Direito, seja qual for a teoria empregada para a compreensão do fenômeno jurídico.

Pode-se reconhecer um papel adicional para a argumentação jurídica, que, acoplado àqueles outros aludidos no item 1, perpassa-os transversalmente, qual seja, o de expor como o aplicador do Direito racionalizou e construiu a atribuição de sentido aos elementos do caso concreto ou abstrato que resulte em sua proposição[71].

Como aponta Ernesto Grün, não é possível saber o que se passa nas cabeças dos juízes ao sentenciar, mas se pode observar a exteriorização desse processo por meio do que dizem em suas sentenças[72].

Acerca dessa temática, Zanon[73] aduz que a racionalidade individual é reciprocamente complementar à racionalidade intersubjetiva, pois ambas

70 Como, por exemplo, já ocorre na Estônia: vide SILVA, Rafael Rodrigues da. *Estônia está desenvolvendo o primeiro "juiz robô" do mundo* (2019). Disponível em: https://canaltech.com.br/inteligencia-artificial/estonia-esta-desenvolvendo-o-primeiro-juiz-robo-do-mundo-136099/. Acesso em: 1º fev. 2021. O próprio STF já sinalizou pela ampliação do uso do Victor para outras atividades, conforme se vê em: BRASIL. Supremo Tribunal Federal. *Inteligência artificial*: trabalho judicial de 40 minutos pode ser feito em 5 segundos (2018); e BRASIL. Supremo Tribunal Federal. *Inteligência artificial vai agilizar a tramitação de processos no STF* (2018). Disponível em: http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=380038. Acesso em: 1º fev. 2021.

71 PERELMAN, Chaim. *Lógica jurídica*. Nova retórica. Trad. Vergínia K. Pupi. São Paulo: Martins Fontes, 2000. p. 222. Título original: *Logique juridique*; MOREIRA, Âmalin Aziz Sant'Ana. *Evolução do conceito de sentença no direito processual civil brasileiro*. Dissertação (Mestrado em Direito) – Universidade Gama Filho. Disponível em: http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=83591. Acesso em: 1º fev. 2021, p. 154.

72 GRÜN, Ernesto. *Una visión sistémica y cibernética del Derecho*. Disponível em: https://www.researchgate.net/profile/Ernesto_Gruen/publication/236151315_UNA_VISION_SISTEMICA_Y_CIBERNETICA_DEL_DERECHO_EN_EL_MUNDO_GLOBALIZADO_DEL_SIGLO_XXI/links/579fb42608ae100d38065b71/UNA-VISION-SISTEMICA-Y-CIBERNETICA-DEL-DERECHO-EN-EL-MUNDO-GLOBALIZADO-DEL-SIGLO-XXI. Acesso em: 1º fev. 2021, p. 59.

73 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 34 e 35.

representam dois momentos coligados e indissociáveis de operacionalização da linguagem: o primeiro hermenêutico e o segundo discursivo e argumentativo. A linguagem serve como meio para interpretar a realidade, raciocinar e transmitir conhecimentos, formando verdades transitórias baseadas em consensos[74].

Na mesma linha, Vinícius Almada Mozetic vislumbra uma unicidade entre compreender, interpretar, explicar e aplicar, sendo todos momentos de um mesmo processo[75].

É de grande importância essa complementaridade, ou até unicidade, pois permite detectar e evitar o arbítrio[76].

A justificação dos posicionamentos dos órgãos jurisdicionais tem, nesse sentido, um papel político, propiciando o controle de sua atuação e a devida responsabilização em caso de irregularidade, o que configura a denominada *accountability* decisional[77].

No item 2, foi apontado que os sistemas de IA dotados da tecnologia de aprendizado profundo são capazes de produzir textos jurídicos a partir de problemas formulados em linguagem natural e, até mesmo, de desempenhar atividades decisórias.

Todavia, se não há dúvida de que os robôs podem elaborar arrazoados contendo argumentos jurídicos, resta perguntar se eles podem efetivamente argumentar.

Mozetic entende que não, pois, para ele, os robôs não são capazes de compreender o mundo, de "abraçar a natureza complexa do raciocínio jurídico"[78]. Faltaria, assim, a primeira dimensão da aplicação do Direito, a hermenêutica, indissociável da proposição de solução para problemas concretos ou abstratos.

---

74 ZANON JÚNIOR, Orlando Luiz. *Teoria complexa do Direito*. Op. cit., p. 34 e 35.

75 MOZETIC, Vinícius Almada. Os sistemas jurídicos inteligentes e o caminho perigoso até a teoria da argumentação de Robert Alexy. *Revista Brasileira de Direito*, Passo Fundo, v. 13, n. 3, p. 437 a 454, set./dez. 2017, p. 444.

76 PERELMAN, Chaim. *Lógica jurídica*. Op. cit., p. 222; MOREIRA, Âmalin Aziz Sant'Ana. Op. cit., p. 154.

77 TOMIO, Fabrício Ricardo de Limas; ROBL FILHO, Ilton Norberto. *Accountability* e independência judiciais: uma análise da competência do Conselho Nacional de Justiça (CNJ). *Revista de Sociologia e Política*, v. 21, n. 45, mar. 2013, p. 29 a 46, p. 30.

78 MOZETIC, Vinícius Almada. Op. cit., p. 444 e 449.

Além disso, a justificação das escolhas é reputada, historicamente, como um dos grandes desafios da IA no Direito[79]. Mecanismos de aprendizado profundo tornam-se mais opacos quanto mais camadas vão sendo acrescentadas às respectivas redes neurais, constituindo verdadeiras "caixas-pretas" complexas, de modo que as razões pelas quais a máquina atingiu determinada conclusão não são acessíveis, nem mesmo a especialistas em computação[80].

Evidentemente, é possível administrar a entrada de dados, escolhendo o conjunto de informações sobre as quais o robô atuará (o *data set*), bem como avaliar as respostas do sistema, executando, a partir delas, sintonias finas nos seus componentes[81]. Esse tipo de ajuste é característico do treinamento das redes neurais e é por meio dele que elas "aprendem". Entretanto, ainda assim, não é possível saber o que exatamente ocorre nas camadas de neurônios artificiais, quando são realizados os ajustes.

Daí que a opacidade do aprendizado profundo torna difícil afirmar que os respectivos robôs efetivamente argumentem.

A argumentação contida nas propostas de decisão ofertadas pelos robôs não revela o raciocínio empregado para construí-las, pois, como visto no item 2, esses sistemas resolvem problemas mediante associações, comparações e reconhecimento de padrões a partir de amplos bancos de dados (*data sets*). A adição de número cada vez maior de neurônios artificiais em interação torna o mecanismo, progressivamente, menos inteligível[82]. As soluções de IA atuais contam com milhões de unidades e bilhões de pesos, de modo que os algoritmos criam uma realidade verdadeiramente complexa, emergente a partir de sua interação[83].

Sejnowski[84] reconhece esse quadro e, diante dele, sugere que a opacidade possa ser resolvida com a intensificação do desenvolvimento das redes neurais, de sorte que os próprios robôs possam efetivamente esclare-

79 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 28.

80 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 99 e 100; SEJNOWSKI, Terrence J. Op. cit., p. 213 a 218.

81 MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Op. cit., p. 224; PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 99 e 100; SEJNOWSKI, Terrence J. Op. cit., p. 213 a 218.

82 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 99 e 100.

83 SEJNOWSKI, Terrence J. Op. cit., p. 130, 212 e 214.

84 SEJNOWSKI, Terrence J. Op. cit., p. 134 e 135.

cer como decidiram determinada questão, quando perguntados. Trata-se, porém, de promessa distante, no atual estágio tecnológico.

Rômulo Soares Valentini[85] pondera que o problema da opacidade também afeta os seres humanos, existindo uma "caixa preta dos juízes".

O autor[86] adota o conceito de heurística para referir-se aos atalhos mentais intuitivos que facilitam a tomada de decisões, embora as sujeitem a falhas cognitivas.

Segundo Daniel Kahneman, "a definição técnica de heurística é um procedimento simples que ajuda a encontrar respostas adequadas, ainda que geralmente imperfeitas, para perguntas difíceis"[87].

De acordo com esse conceito, o processo heurístico dos julgadores não necessariamente pode ser extraído dos argumentos exteriorizados para a solução dos problemas que se apresentam[88].

No entanto, embora o processamento da linguagem em sistemas de aprendizado profundo tenha, realmente, sido inspirado em descobertas neurocientíficas sobre o funcionamento do cérebro humano nessa seara, tais desenvolvimentos são ainda muito incipientes[89].

Ademais, as redes neurais não conseguem, por ora, emular a totalidade das características que constituem a inteligência humana, caracterizada por uma interação complexa de diversos arranjos cognitivos e habilidades comportamentais, resultante em uma capacidade de resolver problemas com engenho, criatividade e até mesmo senso de humor[90].

Nesse contexto, falta mesmo aos robôs a compreensão do mundo de que falava Mozetic, a qual é essencial para a atividade de argumentação.

Cabe questionar, então, se seria o caso de banir a tecnologia do aprendizado profundo das soluções de IA, empregadas na aplicação do Direito.

---

85 VALENTINI, Rômulo Soares. *Julgamento por computadores?* As novas possibilidades da juscibernética no século XXI e suas implicações para o futuro do direito e do trabalho dos juristas. Tese (Doutorado em Direito) – Universidade Federal de Minas Gerais. Disponível em: https://repositorio.ufmg.br/bitstream/1843/BUOS-B5DPSA/1/vers_o_completa_tese_romulo_soares_valentini.pdf. Acesso em: 1º fev. 2021, p. 108 e 109.

86 VALENTINI, Rômulo Soares. Op. cit., p. 43 a 49.

87 KAHNEMAN, Daniel. *Rápido e devagar*: duas formas de pensar. Rio de Janeiro: Objetiva, 2012. p. 127.

88 VALENTINI, Rômulo Soares. Op. cit., p. 108 e 109.

89 SEJNOWSKI, Terrence J. Op. cit., p. 272 a 274.

90 PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. Op. cit., p. 31 e 32.

Quanto a essa dúvida, Valentini[91], por exemplo, entende que deve ser evitado o aprendizado profundo no campo da decisão jurídica, justamente porque, em sua visão, apresenta opacidade semelhante à do cérebro humano. Ele[92] propõe lançar mão de sistemas especializados, erigidos a partir da algoritmização transparente da atividade decisória, sob o modelo da "árvore de decisão", de modo que se possa aferir e controlar cada escolha realizada. O uso desses sistemas corrigiria as opacidades, tanto as computacionais como também as humanas[93].

Ocorre que sistemas como esses, essencialmente baseados na lógica binária, não se beneficiam da plasticidade das redes neurais, as quais, ao balancear o uso de métodos analógicos neuromórficos com os digitais, permitem modular os tempos de ativação de suas camadas conforme várias categorias diferentes de sinais e sinapses, ampliando as possibilidades de computação de informações, com grande economia de energia[94].

Notadamente, as redes exclusivamente baseadas na lógica binária estão sujeitas a limites mais estreitos para sua eficiência, dimensão, velocidade e capacidade de processamento, justamente em face de seu elevado consumo de energia[95]. Até mesmo Valentini admite que o aprendizado profundo apresenta maiores possibilidades, em termos de automação[96].

A proscrição de uma tecnologia também não é uma boa política, especialmente se ela é disruptiva, porquanto prejudica seu desenvolvimento e turba seu potencial de eficiência e redução de custos[97].

Sebastião Tavares Pereira[98], na mesma toada, sustenta que, na relação do direito com as tecnologias digitais, o *approach* deve ser automatizar o máximo e, nesse processo de forçar as fronteiras do automatizável, descobrir o que eventualmente não o seja.

---

91 VALENTINI, Rômulo Soares. Op. cit., p. 109 e 110.

92 VALENTINI, Rômulo Soares. Op. cit., p. 110 a 122.

93 VALENTINI, Rômulo Soares. Op. cit., p. 110.

94 SEJNOWSKI, Terrence J. Op. cit., p. 229 a 236.

95 SEJNOWSKI, Terrence J. Op. cit., p. 229 a 236.

96 VALENTINI, Rômulo Soares. Op. cit., p. 66.

97 Nesse sentido, vide FIORINO, Daniel J. Regulatory innovation and change. In: DURANT, Robert F.; FIORINO, Daniel J.; O'LEARY, Rosemary (Ed.). *Environmental governance reconsidered*. Challenges, choices and opportunities. 2. ed. Cambridge: The MIT Press, 2017. Edição Kindle. Capítulo 9º.

98 PEREIRA, Sebastião Tavares. Processo eletrônico, máxima automação, extraoperabilidade, imaginalização mínima e máximo apoio ao juiz: ciberprocesso. *Revista do Tribunal Regional do Trabalho da 13ª Região*, João Pessoa, v. 16, n. 1, p. 40 a 66, 2009, p. 57.

Cabe então questionar como admitir a atividade decisória pelas máquinas, se, como visto, elas não são capazes de argumentar. Não seria o caso de admitir que cairiam por terra, assim, as essenciais funções desempenhadas por essa atividade na aplicação do Direito, enfocadas no item 1? Se as máquinas dotadas de tecnologia de aprendizado profundo podem produzir textos contendo argumentos para a aplicação do Direito, mas não podem efetivamente argumentar, a atividade argumentativa não parece estar um pouco além do limite do não automatizável, ainda que dentro de algo como uma zona fronteiriça? Haveria alguma abertura nessa fronteira?

Mozetic[99], baseado nas lições de Hans Georg Gadamer, indica um caminho para transitar nessa seara, ao propor a humanização da tecnologia, mediante a sua colocação "a serviço do homem, do sujeito, do intérprete"[100].

Gadamer vislumbra a incorporação das tecnologias informacionais à cultura humanista, em cujo âmbito servirão como instrumentos para implementação dos horizontes de possibilidade de atuação da sociedade[101].

Segundo a visão gadameriana da hermenêutica, o mundo da vida não pode ser experimentado e articulado diretamente por quem pretenda conhecê-lo, mas unicamente por meio de articulações linguísticas, que servem como mediações[102]. A verdade sobre o sentido do mundo e suas conexões será necessariamente uma verdade mediada, ou seja, uma verdade articulada linguisticamente[103], que assim traz consigo toda a carga do contexto histórico no qual foi formulado o problema que ensejou a sua busca[104].

Nessa linha, para que o aprendizado profundo possa se converter em uma mediação apta a proporcionar uma atribuição de sentido ao mundo, ele deve ser suscetível de se articular linguisticamente, ou seja, deve fazer parte da atividade discursiva, inerente à argumentação jurídica. Sendo os seus achados inseridos na argumentação, o aprendizado profundo pode atravessar a passagem entre o automatizável e não automatizável, nova-

99 MOZETIC, Vinícius Almada. Op. cit., p. 445 a 450.

100 MOZETIC, Vinícius Almada. Op. cit., p. 445.

101 FERRERES, José M. Rubio. Hermenéutica y medios de comunicación. In: ACERO, J. J *et al*. (Ed.). *El legado de Gadamer*. Granada: Editorial Universidad de Granada, 2004. p. 519 a 535, p. 532.

102 FERRERES, José M. Rubio. Op. cit., p. 532 e 533.

103 FERRERES, José M. Rubio. Op. cit., p. 532.

104 LIMA, Renata Albuquerque; BRITO, Anya Lima Penha de. Uma análise crítica à luz da hermenêutica dos sistemas jurídicos inteligentes. *Revista Meritum*, Belo Horizonte, v. 14, n. 2, p. 690 a 707, jul./dez. 2019, p. 701.

mente utilizando-se a ideia de Pereira, para o fim de ingressar na intersubjetividade característica do Direito.

Há grandes possibilidades de humanização e inserção dessa tecnologia na cultura jurídica, por meio da argumentação, de modo a servir como instrumento para ampliação das possibilidades de atuação humana, na seara do Direito. Podem ser articulados motivos para acomodar todas as funções da argumentação, tratadas no item 1, quaisquer que sejam os paradigmas nos quais se insiram.

O uso de sistemas de IA para desafogar o Poder Judiciário, produzindo decisões em bloco de matérias idênticas e permitindo aos juízes dedicar-se com maior atenção aos casos mais complexos[105], por exemplo, pode ser considerado um argumento pragmático para a utilização dos robôs em casos como os descritos, pois terá consequências favoráveis.

A preocupação com a performatividade do STF é uma das principais tônicas do Projeto Victor, enfatizando-se a aceleração dos trâmites, a velocidade nos julgamentos, a redução do número de processos acumulados e, ainda, a diminuição de gastos com o Poder Judiciário, concretizando-se, assim, o direito à razoável duração do processo, com menor custo[106]. A IA foi articulada, aqui, em um argumento fundado na melhor aplicação do direito positivo.

Também é possível mencionar a denominada "objetividade algorítmica", ou seja, a confiança dos usuários, no sentido de que os algoritmos sejam ferramentas livres de subjetividade e erro, acurados e intocados por influências e intervenções indevidas[107]. Tal objetividade eliminaria preconceitos humanos do processo de decisão, tornando-a mais justa[108].

Contudo, nesse particular, não se pode esquecer a ressalva de Kahneman, no sentido de que há considerável aversão à tomada de decisões por algoritmos, muito embora existam indicativos de que possam fornecer uma decisão mais precisa, segundo as bases científicas disponíveis, sem os

105 LIMA, Renata Albuquerque; BRITO, Anya Lima Penha de. Op. cit., p. 691.

106 MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Op. cit., p. 221, 222, 226, 227 e 230.

107 GILLESPIE, Tarleton. The relevance of algorithms. In: GILLESPIE, Tarleton; BOCZKOWSKI, Pablo; FOOT, Kirsten (Org.). *Media technologies*. Essays on communication, materiality and society. Cambridge: MIT Press, 2014. p. 167 a 193, p. 179.

108 SOUZA, Cássio Bruno Castro; LEMOS, Vinícius da Silva. Mm. Robot: um devido processo tecnológico para um processo civil automatizado. In: HECKTHEUER, Pedro Abib; LOURENÇO, Bruna Borges Moreira; HECKTHEUER, Márcia Abib (Org.). *Desafios socioambientais das sociedades de consumo, informacional e tecnológica*. 1. ed. Itajaí: Univali, v. 1, 2018. p. 114 a 154, p. 127.

erros intuitivos (heurísticas e vieses) que permeiam o raciocínio humano. Na visão do autor, isso decorre da característica humana de conferir mais valor à causa do erro do que à prevalência técnica e matemática do protocolo de decisão[109]. Em outras palavras, de acordo com suas pesquisas na área da psicologia comportamental, as pessoas preferem ser governadas por seus semelhantes, e não por protocolos.

Por fim, observe-se que não se está manifestando concordância ou discordância com tais argumentos, mas, unicamente, exemplificando como a atividade argumentativa pode abarcar, em seu bojo, elementos atinentes ao uso da IA, sem que seja o próprio robô quem esteja determinando o resultado dessa atividade.

CONSIDERAÇÕES FINAIS

O aprendizado profundo é uma tecnologia disruptiva e promissora. As ideias que a animam apresentam múltiplas facetas de contato com o Direito, especialmente sua racionalidade prática, voltada à solução dos problemas da sociedade. A rapidez, a economia e a precisão das soluções nela baseadas também a tornam atraente para o campo jurídico, assim como as amplas possibilidades de manejo da linguagem que ela apresenta.

No entanto, como qualquer tecnologia, o aprendizado profundo apresenta limitações. Como visto, sua opacidade impede que a produção de arrazoados jurídicos, por parte de robôs, possa ser considerada como efetiva atividade argumentativa, prejudicando, assim, o desempenho automatizado de atividades decisórias, no âmbito jurídico, ao menos no atual estágio tecnológico.

Uma possibilidade, para que não se percam os benefícios, efetivos e potenciais, bem como para que se possa compatibilizá-los com essa limitação, é o reconhecimento do aprendizado profundo como um instrumento para o intérprete do Direito, ao invés de serem os robôs os próprios intérpretes, ao menos nesse primeiro estágio de absorção da tecnologia.

Trata-se de uma solução ao modo gadameriano, no sentido de empregar a produção dos robôs como um elemento adicional na explicitação no

109 KAHNEMAN, Daniel. Op. cit., p. 285: "[...] para a maioria das pessoas, a causa do erro faz a diferença. A história de uma criança morrendo porque um algoritmo cometeu um erro é mais pungente do que a história da mesma tragédia ocorrendo como resultado de erro humano, e a diferença na intensidade emocional é prontamente traduzida em uma preferência moral".

discurso, para o fim de construir a argumentação engendrada pelo aplicador do Direito, de modo a auxiliar e agilizar a solução de um problema que lhe seja apresentado. Desse modo, a AI não atribuirá sentido ao mundo, mas será um dos elementos que integrarão a interpretação.

REFERÊNCIAS

ALEXY, Robert. *Conceito e validade do Direito*. Trad. Gercélia Batista de Oliveira Mendes. São Paulo: Martins Fontes, 2011. Título original: *Begriff und Geltung des Rechtes*.

_____. *Teoria da argumentação jurídica*. Trad. Zilda Hutchinson Schild Silva. São Paulo: Landy, 2001. Título original: *Theorie der Juristische Argumentation*.

AMARAL, Francisco. Racionalidade e sistema no direito civil brasileiro. *Revista de Informação Legislativa*, Brasília, a. 31, n. 121, p. 233 a 243, jan./mar. 1994.

ANDRADE, Mariana Dionísio de; PINTO, Eduardo Régis Girão de Castro; LIMA, Isabel Braga de; GALVÃO, Alex Renan de Souza. Inteligência artificial para o rastreamento de ações com repercussão geral: o Projeto Victor e a realização do princípio da razoável duração do processo. *Revista Eletrônica de Direito Processual – REDP*, a. 14, v. 21, n. 1, p. 312 a 335, jan./abr. 2020.

ATIENZA, Manuel. *As razões do Direito*. Teorias da argumentação jurídica. Trad. Maria Cristina Guimarães Cupertino. São Paulo: Landy, 2003. Título original: *Las razones del Derecho. Teorías de la argumentación jurídica*.

BRASIL. Supremo Tribunal Federal. Inteligência artificial: trabalho judicial de 40 minutos pode ser feito em 5 segundos (2018). Disponível em: http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=393522. Acesso em: 1º fev. 2021.

_____. Supremo Tribunal Federal. Inteligência artificial vai agilizar a tramitação de processos no STF (2018). Disponível em: http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=380038. Acesso em: 1º fev. 2021.

CASTELLS, Manuel. *A sociedade em rede*. Trad. Roneide Venâncio Majer. São Paulo: Paz e Terra, v. I, 2002. Título original: *The rise of the network society*.

DICIONÁRIO MICHAELIS ONLINE. Sinapse. Disponível em: https://michaelis.uol.com.br/busca?r=0&f=0&t=0&palavra=sinapse. Acesso em: 1º fev. 2021.

DWORKIN, Ronald. *Justice for hedgehogs*. Cambridge: Harvard University Press, 2011.

_____. *Levando os direitos a sério*. Trad. Nelson Boeira. São Paulo: Martins Fontes, 2002. Título original: *Taking rights seriously*.

FERRAJOLI, Luigi. *A democracia através dos direitos*. O constitucionalismo garantista como modelo teórico e como projeto político. Trad. Alexander Araújo de Souza e outros. São Paulo: Revista dos Tribunais, 2015.

FERRERES, José M. Rubio. Hermenéutica y medios de comunicación. In: ACERO, J. J. et al. (Ed.). *El legado de Gadamer*. Granada: Editorial Universidad de Granada, p. 519 a 535, 2004.

FIORINO, Daniel J. Regulatory innovation and change. In: DURANT, Robert F.; FIORINO, Daniel J.; O'LEARY, Rosemary (Ed.). *Environmental governance reconsidered*. Challenges, choices and opportunities. 2. ed. Cambridge: The MIT Press, 2017. Edição Kindle. Capítulo 9º.

GILLESPIE, Tarleton. The relevance of algorithms. In: GILLESPIE, Tarleton; BOCZKOWSKI, Pablo; FOOT, Kirsten (Org.). *Media technologies*. Essays on communication, materiality and society. Cambridge: MIT Press, p. 167 a 193, 2014.

GRÜN, Ernesto. Una visión sistémica y cibernética del derecho. Disponível em: https://www.researchgate.net/profile/Ernesto_Gruen/publication/236151315_UNA_VISION_SISTEMICA_Y_CIBERNETICA_DEL_DERECHO_EN_EL_MUNDO_GLOBALIZADO_DEL_SIGLO_XXI/links/579fb42608ae100d38065b71/UNA-VISION-SISTEMICA-Y-CIBERNETICA-DEL-DERECHO-EN-EL-MUNDO-GLOBALIZADO-DEL-SIGLO-XXI. Acesso em: 1º fev. 2021.

KAHNEMAN, Daniel. *Rápido e devagar*: duas formas de pensar. Rio de Janeiro: Objetiva, 2012.

KARAM DE CHUEIRI, Vera. *A filosofia jurídica de Ronald Dworkin como possibilidade de um discurso instituinte de direitos*. Dissertação (Mestrado em Direito) – Universidade Federal de Santa Catarina. Disponível em: https://repositorio.ufsc.br/bitstream/handle/123456789/106357/90511.pdf?sequence=1&isAllowed=y. Acesso em: 1º fev. 2021.

LIMA, Renata Albuquerque; BRITO, Anya Lima Penha de. Uma análise crítica à luz da hermenêutica dos sistemas jurídicos inteligentes. *Revista Meritum*, Belo Horizonte, v. 14, n. 2, p. 690 a 707, jul./dez. 2019.

MAIA FILHO, Mamede Said; JUNQUILHO, Tainá Aguiar. Projeto Victor: perspectivas de aplicação da inteligência artificial ao direito. *Revista Direito e Garantias Fundamentais*, v. 19, n. 3, p. 219 a 238, set./dez. 2018.

MOREIRA, Âmalin Aziz Sant'Ana. *Evolução do conceito de sentença no direito processual civil brasileiro*. Dissertação (Mestrado em Direito) – Universidade Gama Filho. Disponível em: http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=83591. Acesso em: 1º fev. 2021.

MOZETIC, Vinícius Almada. Os sistemas jurídicos inteligentes e o caminho perigoso até a teoria da argumentação de Robert Alexy. *Revista Brasileira de Direito*, Passo Fundo, v. 13, n. 3, p. 437 a 454, set./dez. 2017.

NIEVA FENOLL, Jordi. *Inteligencia artificial y processo judicial*. Madrid: Marcial Pons, 2018.

PASOLD, Cesar Luiz. *Metodologia da pesquisa jurídica* – Teoria e prática. 14. ed. rev., atual. e ampl. Florianópolis: EMais, 2018.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, v. I, 2019.

PEREIRA, Sebastião Tavares. Processo eletrônico, máxima automação, extraoperabilidade, imaginalização mínima e máximo apoio ao juiz: ciberprocesso. *Revista do Tribunal Regional do Trabalho da 13ª Região*, João Pessoa, v. 16, n. 1, p. 40 a 66, 2009.

PERELMAN, Chaim. *Lógica jurídica*. Nova retórica. Trad. Vergínia K. Pupi. São Paulo: Martins Fontes, 2000. Título original: *Logique juridique*.

_____; OLBRECHTS-TYTECA, Lucie. *Tratado da argumentação*. A nova retórica. Trad. Maria Ermantina de Almeida Prado Galvão. São Paulo: Martins Fontes, 2005. Título original: *Traité de l'argumentation*.

POSNER, Richard. *Direito, pragmatismo e democracia*. Trad. Teresa Dias Carneiro. Rio de Janeiro: Forense, 2010. Título original: *Law, pragmatism and democracy*.

SALDANHA, Jânia; MELLO, Rafaela da Cruz; LIMBERGER, Têmis. Do governo por leis à governança por números: breve análise do *trade in service agreement* (TISA). *Revista de Direito Internacional*, v. 13, n. 3, p. 338 a 355, 2016.

SEJNOWSKI, Terrence J. *A revolução do aprendizado profundo*. Trad. Carolina Gaio. Rio de Janeiro: Alta Books, 2019. Título original: *The deep learning revolution*.

SILVA, Rafael Rodrigues da. *Estônia está desenvolvendo o primeiro "juiz robô" do mundo* (2019). Disponível em: https://canaltech.com.br/inteligencia-artificial/estonia-esta-desenvolvendo-o-primeiro-juiz-robo-do-mundo-136099/. Acesso em: 1º fev. 2021.

SOUZA, Cássio Bruno Castro; LEMOS, Vinícius da Silva. Mm. Robot: um devido processo tecnológico para um processo civil automatizado. In: HECKTHEUER, Pedro Abib; LOURENÇO, Bruna Borges Moreira; HECKTHEUER, Márcia Abib (Org.). *Desafios socioambientais das sociedades de consumo, informacional e tecnológica*. 1. ed. Itajaí: Univali, v. 1, 2018.

TEIXEIRA, Mateus. *STF investe em inteligência artificial para dar celeridade a processos* (2018). Disponível em: https://www.jota.info/coberturas-especiais/inova-e-acao/stf-aposta-inteligencia-artificial-celeridade-processos-11122018. Acesso em: 1º fev. 2021.

TOMIO, Fabrício Ricardo de Limas; ROBL FILHO, Ilton Norberto. *Accountability* e independência judiciais: uma análise da competência do Conselho Nacional de Justiça (CNJ). *Revista de Sociologia e Política*, v. 21, n. 45, p. 29 a 46, mar. 2013.

VALENTINI, Rômulo Soares. *Julgamento por computadores?* As novas possibilidades da juscibernética no século XXI e suas implicações para o

futuro do direito e do trabalho dos juristas. Tese (Doutorado em Direito) – Universidade Federal de Minas Gerais. Disponível em: https://repositorio.ufmg.br/bitstream/1843/BUOS-B5DPSA/1/vers_o_completa_tese_romulo_soares_valentini.pdf. Acesso em: 1º fev. 2021.

ZANON JUNIOR, Orlando Luiz. *Teoria complexa do Direito*. 3. ed. rev. e ampl. São Paulo: Tirant lo Blanch, 2019.

_____. Pós-positivismo: a versão pragmática de Posner. *Revista Direito e Liberdade – RDL*, Natal, v. 15, n. 3, p. 117 a 140, set./dez. 2013.

Sobre os autores:

Orlando Luiz Zanon Junior | *E-mail:* olzanon@yahoo.com.br

Juiz de Direito. Doutor em Ciência Jurídica pela Universidade do Vale do Itajaí (Univali). Dupla Titulação de Doutorado em Direito Público pela Università Degli Studi di Perugia (UNIPG). Mestre em Direito pela Universidade Estácio de Sá (UNESA). Pós-Graduação em nível de Especialização pela Universidade do Vale do Itajaí (UNIVALI) e também pela Universidade Federal de Santa Catarina (UFSC). Professor da Escola da Magistratura de Santa Catarina (ESMESC), da Academia Judicial (AJ) e do Programa de Pós-Graduação da UNIVALI. Membro da Academia Catarinense de Letras Jurídicas (ACALEJ). Autor da *Teoria Complexa do Direito* e criador do método de gestão judicial de Triagem Complexa.

Guilherme Kirtschig | *E-mail:* kirtschig2@gmail.com

Procurador do Trabalho. Doutorando em Ciência Jurídica pela Universidade do Vale do Itajaí – UNIVALI, com dupla titulação pela Universidad de Alicante.

Data de submissão: 20 de setembro de 2021.
Data de aceite: 10 de janeiro de 2022.

Todos São Iguais Perante o Algoritmo? Uma Resposta Cultural do Direito à Discriminação Algorítmica

Is Everyone Equal Before the Algorithm? A Cultural Response of Law to Algorithmic Discrimination

ALAN DUARTE[1]
Universidade Federal do Ceará (UFC).

RAMON DE VASCONCELOS NEGÓCIO[2]
Centro Universitário 7 de Setembro (UNI7).

RESUMO: São cada vez mais notórios os casos em que algoritmos de aprendizado de máquina que operam em rede apresentam comportamentos discriminatórios, de modo a evidenciar uma semântica da discriminação que perpassa instituições e estruturas sociais, incorpora-se aos aspectos técnicos da programação e ameaça direitos fundamentais. Evidencia-se, nesse sentido, a importância de se discutir, no âmbito das ciências sociais, sobretudo do Direito, como se construir um modelo de observação, identificação e busca de soluções capaz de lidar com o problema da discriminação perpetrada por algoritmos de aprendizado de máquina, os quais operam majoritariamente por meio da internet. Para solucionar esse problema, o trabalho apresenta, em um primeiro momento, a igualdade compreendida como não discriminação. Questiona-se, em seguida, por meio da análise do funcionamento dos algoritmos, como esses sistemas com capacidade de aprendizado tratam – e como isso ocorre – grupos de maneira desfavorável. Por fim, partindo da compreensão do Direito como uma semântica social, como cultura imbricada nas mais variadas relações sociais, e não apenas nas instituições formais, apresentam-se mecanismos de construção de um modelo de observação, identificação e busca de soluções voltada ao combate da discriminação por algoritmos. A partir de uma pesquisa essencialmente bibliográfica e do método hipotético-dedutivo, conclui-se que a normatividade do Direito está na comunidade de usuários e nas práticas dos programadores, o que revela a importância de impregnar a semântica jurídica da igualdade na seleção e formação de programadores, como possibilidades de solução do problema posto.

1 Orcid: http://orcid.org/0000-0002-0762-1710.
2 Orcid: http://orcid.org/0000-0003-4724-9658.

PALAVRAS-CHAVE: Discriminação por algoritmos; algoritmos de aprendizado de máquina; Direito como cultura.

ABSTRACT: There are increasingly notorious cases in which machine learning algorithms that operate in a network reveal discriminatory behaviors. They demonstrate a semantics of discrimination that permeates social institutions and structures, and end up being incorporated into technical aspects of programming, thus threatening fundamental rights. In this sense, it is fundamental to discuss, in the realm of social sciences, and especially in the realm of law, how to build a model of observation, identification, and search for solutions capable of dealing with the problem of discrimination perpetrated by machine learning algorithms, which operate mostly on the internet. To solve this problem, this paper deals, at first, with the concept of equality in the sense of non-discrimination. By analyzing the functioning of algorithms, I then turn to the question of how these systems with learning capacity treat groups in an unfavorable way, and how this occurs. Finally, taking law as a social semantics, as a culture intertwined with all sorts of social relations, and not only with formal institutions, we present mechanisms to build a model of observation, identification, and search for solutions aimed at combating discrimination by algorithms. From an essentially bibliographic methodology and the hypothetical-deductive method, we conclude that the normativity of law is to be found in the community of users and in the practices of programmers, which reveals the importance of embedding the legal semantics of equality in the selection and training of programmers as one of the possible solutions for the problem here identified.

KEYWORDS: Discrimination by algorithm; machine learning algorithms; Law as culture.

INTRODUÇÃO

Joy Buolamwini, quando era estudante de Ciência da Computação em Georgia Tech, trabalhava com robôs sociais, tendo como um de seus projetos construir um desses robôs para jogar *peek-a-boo*[3]; para a execução dessa tarefa, Joy utilizou um *software* genérico de reconhecimento facial. Ocorre que ela encontrou um problema nessa tarefa: o sistema que ela desenvolveu não conseguia identificar seu rosto, muito embora conseguisse identificar o rosto de sua colega de quarto. A diferença entre ambas e o fator determinante para a falha no reconhecimento é que Joy é negra (Buolamwini, 2016). Esse problema, todavia, não ocorreu apenas de forma isolada com Joy. Repetiu-se em outros contextos e com repercussões mais dramáticas,

3 *Peek-a-boo* é um jogo, comumente jogado com crianças, em que uma pessoa esconde o rosto, principalmente com as mãos, e de repente descobre o rosto e diz "*peek-a-boo*", o que, em tradução livre, seria algo como "te achei".

pois um sistema de reconhecimento facial levou ao etiquetamento de Jacky Alcine e sua amiga, ambos negros, como "gorilas" (Google, 2015), e o uso de outro pelo sistema policial resultou na prisão de Robert William, rapaz também negro, por um delito que ele não cometeu (Hill, 2020).

Situações como essas demonstram que a exclusão de determinados grupos pode e é incorporada por algoritmos de tomada de decisão, de modo que tais sistemas não são neutros ou mais precisos quanto pretendem ou como advogam aqueles que defendem o seu uso irrestrito. Isso porque os algoritmos carregam em sua programação, além de dados brutos e cálculos estatísticos, crenças, vieses e equívocos humanos, pois as escolhas que levam à construção desses sistemas são feitas por seres humanos falíveis[4]. Além disso, os próprios dados que são carregados nesses sistemas são, em certa medida, enviesados, tendenciosos e/ou não representam adequadamente todos os setores da sociedade que deveriam representar.

Isso revela que as formas de discriminação se apresentam de maneiras mais sofisticadas, sendo incorporadas em processos técnicos e perpetuadas por algoritmos de computador, considerados ainda, por muitos, como mecanismos exatos e isentos de subjetividades. Diante desse cenário, os mecanismos jurídicos clássicos, sobretudo as normas de cunho antidiscriminatório, apresentam-se como insuficientes, principalmente por carregarem, em seu bojo, a ideia de intencionalidade na conduta discriminatória e por não compreenderem adequadamente as estruturas socioculturais e a semântica da discriminação que permeia os constructos sociais e técnicos.

A partir disso, o presente trabalho tem como pergunta principal a seguinte: como é possível construir um modelo de observação, identificação e busca de soluções capaz de lidar com o problema da discriminação perpetrada por algoritmos de aprendizado de máquina, os quais operam majoritariamente por meio da internet? Para solucionar esse problema, o trabalho parte de uma pesquisa essencialmente bibliográfica e do método hipotético-dedutivo e divide-se em três partes, além de introdução e conclusão. Em um primeiro momento, apresenta-se a igualdade compreendida como não discriminação, concretamente aduzida e não apenas abstratamente consi-

4 Sobre esse ponto há uma vasta literatura, dentre a qual se destaca: BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. *Calif. L. Rev.*, v. 104, p. 671, 2016; O'NEIL, Cathy. *Weapons of math destruction*: how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016; EUBANKS, Virginia. *Automating inequality*: how high-tech tools profile, police, and punish the poor. St. Martin's Press, 2018.

derada. Em seguida, questiona-se como os sistemas algorítmicos, sobretudo os que operam de forma autônoma, isto é, com capacidade de aprendizado, tratam – e como isso ocorre – grupos de maneira desfavorável. Tal questionamento é feito por meio da análise do funcionamento desses sistemas, bem como pela exploração das vias capazes de fazer o algoritmo redundar em comportamentos discriminatórios. Por fim, a partir da compreensão do Direito como uma semântica social, como cultura imbricada nas mais variadas relações sociais, e não apenas nas instituições formais, apresentam-se mecanismos de construção de um modelo de observação, identificação e busca de soluções voltado ao combate da discriminação perpetrada por sistemas algorítmicos.

IGUALDADE COMO NÃO DISCRIMINAÇÃO

O debate acerca da discriminação algorítmica tem como plano de fundo, do ponto de vista jusfilosófico, a antiga, relevante e profunda discussão sobre a igualdade. Em um primeiro momento, tem-se a igualdade formal, a qual é reduzida à fórmula "todos são iguais perante a lei". Essa ideia de igualdade é adotada e disseminada dentro da primeira geração de direitos, a qual valoriza o homem-singular, de liberdades abstratas (Bonavides, 2017). A compreensão baseada nas ideias do liberalismo clássico (de que todos são iguais perante a lei), embora insuficiente nos dias de hoje, possibilitou a eliminação de privilégios de certos grupos a partir da premissa de que todos possuem uma identidade comum como cidadãos (Moreira, 2017, p. 50). O indivíduo, concebido como uma abstração singular e não corporificada, era a categoria universal do ser humano.

Todavia, essa igualdade formal passa a ser um problema exatamente naquilo que ela apresentava como mais positivo à época de sua instauração, na medida em que se pressupõe que se escolhe um padrão específico de indivíduo ao qual os demais, para serem considerados sujeitos de direitos, deveriam se igualar. Nesse sentido, "essa igualdade formal se corrompe ao eleger como parâmetro pressuposto um sujeito social nada abstrato: masculino, branco, europeu, cristão, heterossexual, burguês e proprietário" (Rios, 2012, p. 172-173). A ideia de igualdade jurídica, portanto, passa a estar mais associada à neutralização das desigualdades no exercício dos direitos do que à igualdade de fato (Neves, 2008, p. 166).

A partir dessa especificação do paradigma de indivíduo detentor de direitos fundamentais, exclui-se desse âmbito de proteção o outro, o diferen-

te. A diferença, portanto, era visibilizada apenas para conceber esse outro como um ser esvaziado de qualquer dignidade, descartável, uma verdadeira *res*, objeto de compra e venda (como na escravidão) ou de extermínio sob o pretexto de purificação do eu (que se contrapõe ao outro), como foi no nazismo (Piovesan, 2008). Nessa perspectiva, destaca Joan Scott (2005, p. 18):

> [...] nesses momentos – quando exclusões são legitimadas por diferenças de grupo, quando hierarquias econômicas e sociais favorecem certos grupos em detrimento de outros, quando um conjunto de características biológicas ou religiosas ou étnicas ou culturais é valorizado em relação a outros – que a tensão entre indivíduos e grupos emerge. Indivíduos para os quais as identidades de grupo eram simplesmente dimensões de uma individualidade multifacetada descobrem-se totalmente determinados por um único elemento: a identidade religiosa, étnica, racial ou de gênero.

Emerge, portanto, em contraposição e complementação a esse cenário, a compreensão de um direito, igualmente fundamental, à diferença, à diversidade, ao lado do direito à igualdade. Isso não se trata de um esforço constitucional por homogeneização social. Muito pelo contrário: o princípio da igualdade busca o respeito entre sistemas sociais de modo recíproco e com simetria às diferenças (Neves, 2008, p. 167; Luhmann, 2016, p. 93). Assim, o princípio da igualdade se revela como uma moeda, cujos lados apresentam a neutralização da desigualdade no acesso ao direito e a manutenção de uma esfera pública plural (Neves, 2008, p. 170).

Assim, enquanto o princípio da igualdade se preocupa em formular uma compreensão do conteúdo e a extensão dessa cláusula constitucional, o conceito de discriminação "aponta para a reprovação jurídica das violações do princípio isonômico, atentando para os prejuízos experimentados pelos destinatários de tratamentos desiguais" (Rios, 2008, p. 19). Ressalta-se, portanto, a compreensão da igualdade material – não como substitutivo, mas como complementação à igualdade formal –, correspondente tanto ao ideal de justiça social e distributiva, quanto ao reconhecimento de identidades, como necessária para que a própria noção de igualdade não se corrompa e justifique ofensas graves à dignidade humana.

Destaque-se, desde já, que as discriminações não são, *per se*, ruins ou proibidas, uma vez que elas, em diversos casos, fazem-se necessárias para a adequada realização do princípio da igualdade. Existe uma assimetria entre a norma de tratamento igual (tratar os iguais de forma igual) e a norma de

tratamento desigual (tratar os desiguais de forma desigual), cuja consequência é a possibilidade de compreender o princípio da igualdade como norma que, *prima facie*, exige um tratamento igual e que permite um tratamento desigual apenas se isso for justificado por princípios contrapostos (Alexy, 2017, p. 411). Busca-se, com as normas antidiscriminatórias, a eliminação de discriminações que mitiguem, eliminem ou prejudiquem, em alguma medida, o exercício de direitos fundamentais de modo injusto.

Sendo assim, o direito fundamental à não discriminação decorre da compreensão do princípio da igualdade, visto sob a perspectiva substancial, e não apenas pela compreensão de uma igualdade formal perante a lei. Essa compreensão substancial destaca uma dimensão concreta do princípio da igualdade – pois concebe os indivíduos dentro de uma realidade social contingente, e não apenas como sujeitos abstratamente considerados aos quais e para os quais deve ser reconhecido o mesmo valor e direcionado um mesmo tratamento –, e se traduz na busca da inclusão social daqueles sujeitos excluídos e subordinados em razão de certas características, como cor, gênero, orientação sexual, dentre outras. Isto é, verifica-se com o direito à não discriminação uma dimensão emancipatória do princípio da igualdade, na medida em que se pretende a eliminação de padrões de exclusão social (Moreira, 2017).

Nesse sentido, o Direito, cuja função vai para além de uma mera estabilização das expectativas normativas, vincula-se a uma pretensão de correção (Alexy, 2007) e, portanto, deve proteger tais características, já que é em razão delas que indivíduos são impedidos de se desenvolverem na vida social[5] (Moreira, 2017, p. 49). Nesse sentido, visualiza-se a igualdade substancial como uma proibição de discriminação. O princípio da igualdade, por conseguinte, só pode ser um princípio jurídico eficaz na medida em que as instituições sociais identificam os processos responsáveis pelas diferentes formas de discriminação[6] (Moreira, 2017).

5 Para maior compreensão do argumento, ver os seguintes casos: Brown *v.* Board of Education (EUA), no qual escolas impediam o acesso de crianças em razão da cor de sua pele, dividindo escolas de pessoas brancas e escolas de pessoas negras. No âmbito brasileiro, destaque-se o caso de Simone André Diniz (posteriormente levado à Comissão Interamericana de Direitos Humanos – OEA), a qual foi impedida de participar de seleção para uma vaga de empregada doméstica em razão da sua cor de pele, pois a divulgação da vaga explicitava que a pessoa a ser empregada deveria, preferencialmente, ser branca.

6 A capacidade que alguém tem para se desenvolver, como argumentam Grant-Thomas e Powell (2006), depende diretamente do acesso à oportunidade na medida em que as oportunidades são produzidas e reguladas como fruto da interação entre instituições e indivíduos – isto é, o modo como as instituições lidam com as contingências naturais e sociais –, de modo a fornecer e negar conjuntamente o acesso com base em cor, gênero, classe social e outros marcadores de diferença social (Rawls, 2008; Grant-Thomas; Powell,

Embora, como afirma Moreira (2017), haja diversas normas jurídicas que pretendem prevenir e/ou punir práticas discriminatórias, ainda persiste uma semântica discriminatória, decorrente, em grande parte, de uma discriminação institucional (Pager; Shepherd, 2008) e, mais profundamente, de uma discriminação estrutural (Almeida, 2020). Nas palavras de Barocas e Selbst (2016, p. 673-674 – tradução livre), "[i]nconscientes, preconceitos implícitos e inércia nas instituições da sociedade, em vez de escolhas intencionais, são responsáveis por grande parte dos efeitos discriminatórios observados".

Além disso, muitas dessas normas partem de uma compreensão da discriminação apenas como discriminação direta, isto é, implicam os conceitos de intencionalidade e arbitrariedade (Moreira, 2017, p. 17). Todavia, é necessária também uma perspectiva institucional e estrutural da discriminação, uma vez que se preocupa com a gênese da discriminação não intencional levadas a cabo por indivíduos, grupos e organizações – fenômeno inadequadamente respondido por abordagens que enfatizam a intencionalidade como fator determinante para a constatação de práticas discriminatórias, denominadas como discriminação indireta (Rios, 2008). A ideia de discriminação engloba como prática qualquer distinção, exclusão, restrição ou preferência, cuja finalidade seja impedir que os indivíduos desfrutem, em pé de igualdade, de direitos fundamentais, ainda que referidas ocorram despidas de intencionalidade. Ou seja, a discriminação está não só na causalidade, mas também nas correlações (Tischbirek, 2019).

Em uma perspectiva do princípio da igualdade, pode-se entender que, quanto mais se sedimenta a discriminação social negativa – historicamente obstáculo ao reconhecimento de determinados direitos a certos grupos sociais –, mais se valida a discriminação positiva de direitos, orientando o princípio da igualdade à realidade social (Neves, 2008, p. 174). Não é só, todavia, um compromisso de o Estado Democrático de Direito incluir pelo princípio da igualdade. A esfera pública – com a mudança dos paradigmas dos meios de comunicação, isto é, tornando a internet o principal meio de comunicação da sociedade – passa a ser mediada não somente pela imprensa tradicional como também por atores privados que administram intermediários e aplicativos da internet, o que os vincula a um compromisso de

2006). Em razão disso, é preciso "pensar a igualdade jurídica como a igualação jurídica [...] a igualdade posta em movimento, em processo de realização permanente; a igualdade provocada pelo Direito segundo um sentido próprio a ela atribuído pela sociedade" (Rocha, 1996, p. 287).

inclusão pela administração de sua normatividade técnica[7]. Assim, a garantia da igualdade pela heterogeneidade é um compromisso mais amplo que inclui e vincula esses agentes privados e sua programação de algoritmos.

A (DES)IGUALDADE POR ALGORITMOS E PELA INTELIGÊNCIA ARTIFICIAL

A busca por um termo por meio de um motor de busca, a indicação de amigos em redes sociais e o aparecimento de notícias direcionadas ao interesse do usuário são típicos resultados de um processo algorítmico. O termo algoritmo tem entrado crescentemente em análises sociais, especialmente políticas e jurídicas, seja pelos crescentes usos de serviços de busca na internet, seja pelos riscos da perda de privacidade e pelo crescente poder de informação de atores privados. Sob o risco de se tornar um termo performativo para os juristas, os algoritmos precisam de uma definição, quando não, ao menos o entendimento de sua função na internet. Quando se trata de algoritmos neste artigo, refere-se a "um procedimento de tomada de decisão e tecnologia da informação digital", de modo que se diferenciam algoritmos de "avaliação dos dados, geração de conhecimento e, portanto, determinação dos critérios de decisão e os critérios para a decisão em si" (Ernst, 2017, p. 1026 – tradução livre). A partir dessa definição, é possível entender que os algoritmos são um sistema de execução de uma programação de código digital, que pode ser determinada por homens e mulheres, assim como por inteligência artificial, isto é, independente de uma programação humana (Hoffmann-Riem, 2017, p. 3-4). Por sua forma adaptativa, os algoritmos contribuem na individualização e precisão em sistemas de busca e sugestão de decisões a partir dos dados indicados e pelo histórico existente do usuário[8].

Algoritmos "são mecanismos de solução de problemas", sendo "essencialmente definida pela atribuição automatizada de relevância a certas informações selecionadas" (Just; Latzer, 2016, p. 2). Há vários tipos de algoritmos, tais que vão desde os mais conhecidos mecanismos de busca, como o Google Search, passando pelos de produção de conteúdo e de produção

7 Mais à frente será tratada a normatividade como marcação positiva de uma possibilidade. Nesse caso, a normatividade técnica está demarcada positivamente por uma tipificação de padrões técnicos, normas-artefatos, indicadores de comportamento etc. Para mais detalhes, confira: ARTOSI, Alberto. Technical normativity. In: *Italian Philosophy of Technology*, Springer, Cham, p. 149-160, 2021.

8 Tamanha é a influência dos algoritmos, que, segundo Luciano Floridi (2015), não vivemos em um mundo *online* ou *offline*, mas sim *Onlife*. Isso significa que não estão claros os limites do que está ou não no mundo digital e fora dele.

no âmbito jornalístico (ex.: Quill e Quakebot), até os de recomendações (ex.: Amazon e Netflix). Independentemente da finalidade oferecida pelos programas, eles "geram, coletam, processam e agregam numerosos dados por meio de métodos algorítmicos de modo a extrair valores econômicos disso" (Just, 2018, p. 3). Assim, as informações, os comportamentos e as comunicações se individualizam conforme o uso do usuário.

A solicitação do usuário mais as características disponíveis dele na plataforma fazem os algoritmos selecionar elementos de um conjunto de dados básico (*Big Data Set 1*), classificando e gerando relevância contextualizada sob uma avaliação estatística automatizada: o modelo *input-output* de uso dos algoritmos possui três fases: (1) a inserção de dados (*input*), (2) o tratamento e (3) a emissão de dados (*output*). Ao lado da inserção de uma requisição por parte do usuário, o programa já possui uma base de dados que servirá para auxiliar na precisão máxima do requisitado com base em outros dados do usuário. Todo resultado retroalimenta o sistema para uma nova busca: a base de dados do usuário e a requisição (1) servirá de fonte para a seleção do algoritmo de escolha que encaminhará a uma base de dados reduzida, que seguirá para o algoritmo de classificação (2). Este encaminhará a base de dados de saída, que, por fim, servirá para a utilização dos dados de emissão (3)[9].

O que foi explicado acima é uma forma geral da lógica de tratamento de dados por meio de algoritmos. Nem todo tratamento de dados, porém, se refere à inteligência artificial (IA). A despeito de o conceito de inteligência artificial ter se popularizado nos últimos anos, principalmente em razão do avanço e dos usos no dia a dia, seu conceito, ou pelo menos a ideia por trás, é mais antiga e discutida na literatura por várias décadas (Russel; Norvig, 2016), alternando momentos de grande progresso e euforia com momentos de frustrações e esquecimentos (Maini; Sabri, 2017). As pesquisas sobre esse assunto, conforme a maioria dos estudiosos, são inauguradas pelo artigo *Computing machinery and intelligence*, publicado na revista *Mind* em outubro de 1950 pelo matemático britânico Alan Turing, considerado o pai da computação moderna.

9 É possível encontrar um quadro explicativo do funcionamento desse sistema em (Schulz; Dankert, 2016, p. 65); para exemplos, cf. GRIMMELMANN, James. The structure of search engine law. *93 Iowa L. Rev.*, 1 2007-2008, s. 7-11; SAURWEIN, Florian; JUST, Natascha; LATZER, Michael (2017). Algorithmische Selektion im Internet: Risiken und Governance automatisierter Auswahlprozesse. *Kommunikation@gesellschaft*, 18:online., s. 1-2.

O termo inteligência artificial é cunhado pelo *Dartmouth Summer Research Project on Artificial Intelligence*, organizado por, dentre outros, J. McCarthy, M. L. Minsky e C. Shannon (1955). Com isso, a IA passa a ser um campo de estudo da Ciência da Computação que parte da ideia de que os aspectos da inteligência humana, sobretudo o aprendizado e a resolução de problemas, podem ser precisamente descritos ao ponto de uma máquina ser capaz de simular tais características, tendo como problema central "fazer uma máquina se comportar de maneiras que poderiam ser chamadas de inteligentes se um ser humano assim se comportasse" (McCarthy *et al.*, 1955, p. 11 – tradução livre)[10]. Enquanto aplicação prática, IA pode ser definida como o conjunto de técnicas, métodos e sistemas que objetivam fazer um computador exibir comportamentos inteligentes, sobretudo o aprendizado, a partir da coleta e interpretação de dados externos, a fim de atingir objetivos e tarefas específicas por meio de uma adaptação flexível (Kaplan; Haenlein, 2018).

Dentre as diversas técnicas abarcadas pelo conceito de IA, aquela que é responsável por uma grande parcela dos avanços desse campo nos últimos e também aquela em que são concentrados os maiores esforços (Cortiz, 2020, *online*) é a área de *machine learning*. O desenvolvimento e aperfeiçoamento dos processadores, a crescente disponibilidade de dados[11] e a possibilidade de coletá-los e armazená-los, associada ao desenvolvimento tecnológico e maior uso da internet (Goldschmidt; Passos, 2005), permitiram que muitas pesquisas de cunho teórico desenvolvidas desde a década de 1950 (Peixoto; Silva, 2019, p. 87) fossem postas em prática, mediante o desenvolvimento de algoritmos de aprendizado de máquina.

Kevin Murphy define *machine learning* (2012, p. 1) como conjunto de métodos que podem automaticamente detectar padrões em dados, e então usá-los para prever dados futuros ou desempenhar outras formas de tomada de decisão. Em razão disso, os mecanismos decisórios deixaram de ser exclusividade de seres humanos e passaram a ser delegados aos algoritmos aprendizes, uma vez que eles são mais eficientes e eficazes que os humanos em determinadas atividades, como análise de dados e tarefas preditivas. Os algoritmos inteligentes, desenvolvidos pela área de *machine*

10 O termo, contudo, sofre críticas, tais como as de Elena Espósito (2017), que, sob uma perspectiva da teoria dos sistemas, trata o termo como "comunicação artificial".

11 Estima-se que são gerados cerca de 2,5 quintilhões de *bytes* por dia, e essa quantidade tende a aumentar ainda mais (Marr, 2018).

learning, adquirem certo grau de autonomia, na medida em que aperfeiçoa sua *performance* em determinada tarefa, ou seja, aprendem (Mitchell, 1997, p. 3)[12].

Entretanto, a despeito desses benefícios, durante a construção, a programação e o treinamento de algoritmos, é possível que o processo apresente resultados desproporcionalmente adversos, concentrados em grupos historicamente desfavorecidos[13] (Barocas; Selbst, 2016, p. 673). Esses algoritmos apresentam essas tendências, ainda que não sejam programados para isso, tendo em vista a capacidade de aprendizado autônomo desses sistemas. Como ressaltam Barocas e Selbst (2016), a discriminação pode ser um artefato próprio do processo de aprendizagem dos algoritmos, em vez de um resultado decorrente do fato de os desenvolvedores atribuírem a determinados fatores pesos inadequados. Isso revela que a exclusão de determinados grupos pode e é incorporada pelo processo técnico que constitui a programação algorítmica, de modo que tais sistemas não são neutros ou mais precisos o quanto pretendem ou como advogam aqueles que defendem o seu uso irrestrito.

A programação de algoritmos não inteligentes pode ser direcionada propositalmente para a discriminação de grupos ou possuir algum processo negligente em não adequar seus dados para evitar efeitos discriminatórios. É na IA, todavia, que reside maior complexidade sobre o tema. Solon Barocas e Andrew D. Selbst (2016) elencam alguns mecanismos da construção técnica dos algoritmos de IA que podem redundar em efeitos discriminatórios. O primeiro diz respeito ao desenho do modelo (ou *design*). Com o advento e a modernização das técnicas de *machine learning*, tornou-se possível a criação de algoritmos aprendizes, algoritmos que não precisam de instruções completas e exaustivas para que ele possa executar uma tarefa, pois eles apresentam uma capacidade de aprendizado autônomo a partir dos dados externos. Esses algoritmos, conforme explica Pedro Domingos (2017),

12 Em razão disso, o setor privado utiliza massivamente esses mecanismos visando, pelo menos, a quatro finalidades mercadológicas principais: i) previsibilidade e diminuição de riscos inerentes à atividade empresarial; ii) interação direta com o consumidor, de modo a prover as necessidades individuais de cada um e assim conseguir captar e fidelizar a maior quantidade de clientes possível; iii) diferenciação de produtos e iv) de serviços, diferente do modelo anterior, que direcionava o mesmo produto/serviço a várias pessoas (Mendes, 2014).

13 Exemplos claros disso são os casos destacados por Ryan Calo (2017, p. 9): um sistema de tradução instantânea que associa a profissão de engenheiro como masculino e a de enfermeira como feminino (mesmo que o idioma traduzido não faça distinção de gênero) e uma câmera que se recusou a fotografar uma blogueira de Taiwan por "acreditar" que ela estava de olhos fechados.

criam novos algoritmos, os modelos. Mas, para criar esse modelo, é preciso que seja especificado para o computador, na forma de um algoritmo de treinamento, o que ele precisa aprender e a partir do que ele irá aprender. Os desenvolvedores, portanto, devem traduzir um certo problema amorfo em uma pergunta que possa ser expressa em termos mais formais de modo que os computadores possam analisar (Barocas; Selbst, 2016, p. 678). Essa tarefa de tradução é essencialmente subjetiva, na medida em que exige a representação, em linguagem que o computador possa compreender, de um determinado problema.

Se o desenvolvedor rotula um conjunto de dados, os quais são usados para treinar o modelo, como "bom pagador" e outro como "mau pagador" e faz isso com base em critérios próprios, ou seja, naquilo que ele acredita ser a definição mais adequada de "bom/mau pagador", ele estará expondo ao modelo um conjunto de dados cujos rótulos refletem uma perspectiva específica e subjetiva sobre um fenômeno maior e o sistema irá inferir desse conjunto de dados um certo padrão, o qual será aplicado para outros dados, distintos daqueles do *dataset training*. Essas novas classes criadas pelo desenvolvedor apresentam alguns pontos cegos (O'Neil, 2016), em razão da limitação epistêmica – ou seja, a definição do que é bom adotada refletirá apenas um aspecto dentro de um amplo espectro de possibilidades –, os quais poderão ser irrelevantes ou, ao contrário, apresentar efeitos nefastos.

O segundo problema diz respeito à seleção/tratamento dos dados utilizados pelos algoritmos de treinamento para treinarem os modelos que serão exportados. A maioria dos problemas em que são utilizados algoritmos aprendizes depende, para a adequada solução do problema, de uma grande base de dados da qual o modelo irá extrair os padrões acerca daquele fenômeno objeto da previsão, isto é, os exemplos por meio dos quais o modelo será treinado para agir de determinada maneira. Como destacam os cientistas da computação: dados tendenciosos (ou enviesados) levam a modelos discriminatórios (Custers, 2013).

Se a aprendizagem de máquina trata determinados casos nos quais o preconceito desempenhou algum papel como exemplos válidos para aprender – por exemplo, se o banco de dados utilizado foi construído com base em decisões humanas diretamente influenciadas por preconceitos, como a contratação mais frequente de homens brancos e alta rotatividade de pessoas pertencentes a grupos minoritários –, essa regra pode simplesmente reproduzir o preconceito envolvido nesses casos anteriores (Barocas; Selbst, 2016, p. 680-681). Em poucas palavras, os dados refletem situações passa-

das e, ao serem incorporados nos algoritmos de tomada de decisão, definirão comportamentos futuros.

Há ainda, conforme os autores, um outro problema decorrente do uso de dados enviesados: se o processo de aprendizagem de máquina extrai certas inferências de uma amostra tendenciosa da população, qualquer decisão que se apoie nessas inferências pode sistematicamente prejudicar aqueles que estão sub ou super-representados (ou sub- ou sobreintegrado[14]) no conjunto de dados. O problema aqui reside no processo de coleta de dados, ou seja, não decorre das intenções que motivaram a tomada de decisão e que gerou aquele *dataset*, mas a capacidade que aquela base de dados tem de representar adequadamente determinados grupos sociais. No primeiro caso, tem-se um banco de dados alimentado por dados tendenciosos, ou seja, que foram gerados com base em preconceitos; ao passo que, no segundo, há um *dataset* com sub ou super-representação de certos grupos, o qual, de igual modo, pode conduzir a tomada de decisões discriminatórias.

Jonas Lerman (2013), nesse sentido, revela uma preocupação particular acerca do fenômeno chamado por ele de *big data's exclusion*, consistente na (sub)representação de certos grupos nos *datasets* com base nos quais grandes empresas e setores governamentais tomam decisões estratégicas por meio de algoritmos de aprendizagem de máquinas. A preocupação refere-se "a omissão sistêmica e não aleatória de pessoas que vivem às margens do *big data*, seja devido à pobreza, geografia ou estilo de vida, e cujas vidas são menos 'datadas' do que as da população em geral" (Lerman, 2013, p. 57 – tradução livre). A marginalização em setores-chave corre o risco de distorcer conjuntos de dados e, por conseguinte, distorcer as decisões tomadas com base naqueles dados.

Por fim, esses algoritmos se utilizam e, em certa medida, dependem de *proxies*. Uma variável *proxy* descreve algo que provavelmente

14 O termo se encaixa na ideia de Marcelo Neves a respeito da ausência de cidadania diante de exclusão social generalizada (ou ausência de acesso aos direitos disponíveis na ordem jurídica). A exclusão social não permite que pessoas estejam em condições de igualdade legal, isto é, ou estão acima da lei, como os sobreintegrados (conhecem os benefícios, mas não cumprem seus deveres legais) e os subintegrados (conhecem só a esfera punitiva da lei, mas não os seus direitos fundamentais). Muito embora as terminologias usadas não guardem perfeita correspondência, elas convergem na ausência de critérios justos e adequados; no caso da representação de dados em um *dataset training*, as razões pelas quais determinados grupos são mal representados decorrem de uma construção desigual, levada a cabo ao longo dos anos. Essa mesma construção social determina e cria as classificações apresentadas por Neves, isto é, grupos privilegiados que se desigualam por receberem os benefícios da lei, mas desconhecerem os deveres, e grupos que só conhecem os deveres (às vezes, aplicados de maneira desproporcional), mas não são destinatários dos benefícios. Vide Neves, 1994.

não é, por si só, muito interessante para o sistema; entretanto, a partir dela, pode se chegar a uma variável relevante e necessária para o algoritmo (Calders; Žliobaitė, 2013, p. 52-53). Como explica O'Neil (2016), às vezes o algoritmo não terá disponível determinado dado que seja de extrema relevância para o aprendizado de máquina; todavia, esses dados podem ser substituídos pelas variáveis *proxies*, as quais estabelecem correlações estatísticas com as variáveis de interesse.

Dessa forma, por mais que dados relativos aos critérios de discriminação proibidos, como cor, gênero, orientação sexual, dentre outros, sejam "ocultados" do sistema, isto é, não sejam utilizados como atributos para o treinamento do modelo, o algoritmo, exatamente em razão dessas variáveis, poderá inferir estatisticamente que um determinado atributo mantém forte correlação com outro ao ponto de tal correlação ser crucial para o desempenho do sistema. Esse problema deriva daquilo que se chama de "codificações redundantes", que se refere ao caso em que a associação a um grupo minoritário passa a ser codificada em outros dados (Barocas; Selbst, 2016).

Se, por exemplo, um determinado critério de discriminação proibido, como cor ou gênero, possuir grande correlação com algum atributo, com alguma característica que se pretende buscar ou prever, é uma consequência natural que membros daqueles grupos raciais ou de gênero sejam afetados de maneira desproporcional. Utilize-se, para exemplificar esse ponto, o caso em que a contratação de pessoas é feita por algoritmos de aprendizado de máquina e um dos critérios utilizados para selecionar o melhor candidato seja o nível de escolaridade e a instituição de ensino que o candidato frequentou. Caso as melhores instituições (assim consideradas pelo desenvolvedor do sistema) possuam um elevado número de estudantes brancos e homens, então ocorrerá uma codificação redundante, ou seja, a cor e o gênero serão codificados dentro do critério escolaridade.

Percebe-se que os algoritmos e seus dados carregam não só preconceitos, mas também normatividades sociais, o que inclui o próprio direito. Os preconceitos "encapsulados" nos dados estão na dinâmica cultural de um grupo social. O direito também está na cultura. Para se entender o lugar do direito nessa (tensa) relação com os dados enviesados, precisa-se compreender o direito para além da sua noção institucionalizada: o direito como semântica social ou como cultura.

DIREITO COMO CULTURA

O uso de um programa executado por algoritmos produz uma relação colaborativa entre programação e usuários, o que permite a adaptação e melhora da experiência de uso do aplicativo ou da plataforma. Mesmo com a programação para o controle automático, ainda há espaço de participação das comunidades, que é importante para melhorar a adequação da plataforma: avaliação e reputação dos usuários e aplicativos, por exemplo, que contribuem para o melhor controle da programação e – logo – das regras estabelecidas na plataforma[15]. A comunidade serve como uma espécie de "*Gedächtnisorganisation*"[16], em que mudanças aparecem por meio de sugestões ou reclamações, consolidação de padrões técnicos e regras sociais de condutas naquele aplicativo.

Há uma linha tênue entre o aspecto binário e técnico de um algoritmo e as normas sociais. Essa linha é determinada pela programação que trabalha com códigos técnicos, que não se reduzem somente a estabilizar expectativas sociais, mas também garantem sua variabilidade e flexibilidade de conexões entre diferentes tecnologias (Vesting, 2004, p. 664). A resposta para problemas, assim, não é meramente por um processo regulatório de estabilização de expectativas normativas, mas adaptativa pelas indagações de usuários.

A programação passa, portanto, por uma influência da dimensão cultural do Direito: o código técnico não é isolado ou neutro, pois interage com outras estruturas normativas. O olhar da comunidade de usuários e do programador são fundamentais na construção de uma normatividade que está em torno da função de uso do programa. A observação do direito não parte de uma exclusiva e imediata incidência de alguma ordem jurídica localizada. Há, ao contrário, um controle próprio que incorpora diversos conteúdos jurídicos, observando diversas ordens jurídicas e as interações de sua comunidade de usuários. Tal controle passa por uma percepção própria do direito que não se encontra na diferenciação funcional e de papéis e na normatividade sistêmica. Mais que isso: o direito faz parte de uma semân-

15 Sobre comunidades *online*, veja AURAY, Nicolas. Online communities and governance mechanisms. In: BROUSSEAU, Eric; MARZOUKI, Meryem; MÉADEL, Cécile (Hrsg.). *Governance, regulations and powers on the internet*. Cambridge University Press: Cambridge, 2012, ss. 211-231.

16 Em sentido semelhante, LANGE, Andreas. Die Gaming-Community als Pionier der digitalen Bewahrung. In: KLIMPEL, Paul; KEIPER, Jürgen (Hrsg.). *Was bleibt?* Nachhaltigkeit der Kultur in der digitalen Welt. Berlin: iRigts.Media, 2013, ss. 110-117.

tica social dominada também por atores externos de suas práticas formais (Rosen, 2006, p. 6-7).

Os direitos tradicionais, antes concentrados na normatividade das ordens jurídicas estatais, não se encontram somente diluídos em outras ordens jurídicas (Neves, 2009; Teubner, 2003), mas também como algo que está na semântica de práticas sociais, isto é, na cultura, que é incorporada por determinadas práticas técnicas. Portanto, percebe-se que a presença de conteúdos jurídicos normativos, cuja lógica estrutura e organiza relações mapeadas (Ladeur, 2016, p. 284), apresenta-se como elemento da cultura[17].

A cultura é um sistema de imunidade ou identidade de grupos, que querem se orientar em valores comuns, experiências, expectativas e interpretações, que forma o sentido simbólico, a imagem do mundo de uma sociedade (Assmann, 1992, p. 140). Segundo Vesting (2015, p. 21), "[a] cultura então também inclui a ordem simbólica e imaginária que determina como as pessoas percebem, sentem, falam, pensam e agem no mundo e dão sentido e significado à vida"[18]. Mesmo com especialização intensa, o direito é também elemento cultural, pois não é estrutura isolada da sociedade (Rosen, 2006, p. 4-5)[19]. O direito faz parte da imaginação social, que ajuda a compreender o mundo (Rosen, 2006, p. 12). Enquanto cultura, nota-se que o direito é conhecimento local – incluindo plataformas, aplicativos e programas mediados por algoritmos em geral –, como "caracterização vernacular" conectado a "convencimentos vernaculares", que se caracterizariam como "sensibilidade legal" (Geertz, 1983, p. 215). Se, em torno do uso dos programas mediados por algoritmos, existem regras, comunidade de usuário e programador, pode-se entender que o programa é um "lugar", pois isso possibilita visualizar a presença do direito como cultura em comunidades que não possuem um lugar concreto ou um longo histórico de existência.

17 O trabalho pretende colocar o Direito também como prática social. Há trabalhos diferentes (cf. NADER, Laura (Ed.). *Law in culture and society*: with a new preface. Univ. of California Press, 1997) que colocam as mudanças conceituais da noção de Direito em diferentes contextos. Essa noção de cultura e Direito não impede um enfoque diferente: ao invés de Direito *na* cultura, tratamos de Direito *como* elemento da cultura, de modo que ele pode ser operacionalizado fora das instâncias tradicionais e dos papéis sociais diferenciados (juiz, promotor, advogados etc.). Assim, a percepção social de quem pode decidir, crendo normatizar com valor jurídico, torna-se importante, tal como um usuário, programador ou controlador de dados ao interpretar "liberdade de expressão", por exemplo.

18 No mesmo sentido, GEERTZ, C. Dichte Beschreibung, 1983, 46; K.-H. Kohl, Ethnologie, 2000, 167.

19 Em sentido semelhante, GEERTZ, Clifford. *The interpretation of cultures*. New York: Basic Books, 1973, s. 5-6.

A SEMÂNTICA JURÍDICA NA COMUNIDADE DE USUÁRIOS E NO PROGRAMADOR: UMA RESPOSTA PARA O PROBLEMA

As conexões entre saber técnico e normatividade jurídica não seguem uma lógica hierárquica, mas se constroem numa rede heterárquica, que permite novas conexões (Ladeur, 2016, p. 25). Assim, não só as ordens jurídicas por meio de suas normas recebem o saber técnico: este também absorve a normatividade do direito. No contexto de transformação digital, verifica-se "um recuo do direito estabelecido pelo Estado como meio de estruturar situações da vida" e, ao mesmo tempo, "o deslocamento da responsabilidade para portadores privados nas esferas determinadas pela digitalização" (Hoffmann-Riem, 2019, p. 18). Isso permite que apareçam novos atores cognitivos, que agem como marcadores do saber e completam as formas de autocoordenação espontânea (Ladeur, 2016, p. 38). Normas e práticas não estatais coordenam cada vez mais atores da rede, sendo o direito e suas normas dissolvidas sob a descrição da elaboração de padrões técnicos (Ladeur, 2016, p. 40). O direito na amplitude de sua textura aberta permite a capacidade de adequação sob as condições de transformações contínuas da sociedade, que pode observar a ligação das instituições e dos institutos jurídicos e suas formas sociais de coordenação (Ladeur, 2016, p. 55). Desse modo, mais do que um senso de direitos e deveres, o direito se torna uma condicionante e estimulante para que algo novo possa ser produzido na prática (Ladeur, 2016, p. 56).

Uma capacidade de aprendizado e flexibilidade – tanto de ordens jurídicas quanto de práticas técnicas (incluso a programação de algoritmos) – é influenciada por conexões horizontais e de compatibilização de expectativas (Ladeur, 1990, p. 142-143). Não é por outro motivo que se precisa admitir a coexistência do direito com outros "códigos" na internet: "No ciberespaço, devemos entender como um 'código' diferente regula: como o *software* e o *hardware* (ou seja, o 'código' do ciberespaço) que faz do ciberespaço o que ele é, também regula o ciberespaço como ele é" (Lessig, 2006, p. 5 – tradução livre). Contudo, esses códigos técnicos vão observar ao seu modo o normativo do direito. Conteúdos como "pornografia infantil", "desrespeito à privacidade" e "discursos de ódio" são, em vários momentos, incorporados na programação como uma observação geral de várias ordens jurídicas, como se a própria programação antecipasse problemas e os evitasse simultaneamente em uma pré-seletividade. Não há uma negação, portanto, da normatividade jurídica, mas um entrelaçamento

entre esta e os códigos de programação, a partir do que a programação lê do conteúdo normativo.

Neste trabalho, o sentido de norma é entendido em sua extensão contrafactual, isto é, como demarcação positiva de uma possibilidade (Möllers, 2015, p. 125). Toda norma possui uma significação autoral que a demarca como símbolo de algo ou alguém (autoridade). Ela não pode se identificar com a realidade, pois assim perderia a condição de possibilidade, já que sempre seria realizada. Não pode também não possuir uma dimensão completamente irrealizável, dado que não a confirmaria também como possibilidade. Ou seja, o cumprimento e o desvio compõem a noção de normatividade. "A *norma* pode ter sua dimensão textual, representando a autorização normal [e a formalização]" (Möllers, 2015, p. 286), que está associada à *aplicação*. Esta é a realização concreta da própria norma (no caso em espécie), criando outra norma (Möllers, 2015, p. 182-183). Tanto na dimensão textual quanto na aplicação, há a normatividade, ao contrário da execução, que é "uma causa que permite apenas um efeito" (Möllers, 2015, p. 183), não tendo a possibilidade do desvio.

No caso da normatividade do uso de algoritmos, os termos de uso e códigos de conduta de um programa têm uma função de *pré-seletividade* na programação de algoritmos. Pode ser feito um paralelo com o texto legal: a pré-seletividade diminui o campo de possibilidades de ação, mas que pode ou não ser cumprido pelo *uso* do programa. Alguém que, mesmo sabendo que é proibido publicar vídeo com nudez em uma plataforma, pode utilizar de meios não captáveis pelo algoritmo programado e desviar dos termos de uso daquela plataforma. O uso se aproxima da ideia de aplicação da norma, pois, ainda que não perceba, o usuário faz um uso – em termos concretos – do programa carregado de sua pré-seletividade. Ambos, assim, são demarcados positivamente por uma possibilidade.

Por outro lado, a execução dos algoritmos em si tem a função de apenas de obedecer a um comando ou uso dado pelo usuário, que já é filtrado pelas normas de conduta e pela programação feita pelo responsável, não havendo, assim, a possibilidade de desvio ou frustração do caminho normativo. Não há normatividade, portanto, exclusivamente no funcionamento do algoritmo, pois não resta para ele a possibilidade de cumprir ou descumprir, mas apenas seguir a instrução dada.

Por essa estrutura de descrição normativa, o algoritmo pode discriminar, mas também pode incluir. Entendendo que a ponte para uma constru-

ção inclusiva passa pelo uso do programa pela comunidade de usuários e pelo programador, fica mais fácil entender o lugar de como o direito na sua função semântica pode atuar. Os usuários servem como atores de atualização da programação algorítmica: ao notar um resultado discriminatório, os usuários podem contextualizá-lo, denunciando à administração do programa para corrigir o filtro dos dados. Exemplo disso é o caso recente, ocorrido em setembro de 2020, na rede social Twitter, em que se percebeu um comportamento discriminatório na apresentação de imagens. O algoritmo dessa rede parecia "ter uma preferência" por pessoas brancas em detrimento de pessoas negras, uma vez que, independentemente do posicionamento das pessoas numa foto, o algoritmo recortava e apresentava, como pré-exibição da imagem, a parte em que estava a pessoa branca. Em nota, o porta-voz da rede social informou que iria continuar com mais análises para remover do algoritmo vieses raciais, além de abrir o código-fonte para outras pessoas revisá-lo e replicá-lo[20] (Hern, 2020, *online*).

Isso evidencia uma exigência que a programação observe processos de autorregulação, os quais precisam ser claros de modo que possam ser controlados por seus usuários. Essa limitação parece ser influenciada por uma dimensão cultural do direito, uma vez que a codificação técnica não se encontra isolada do meio em que se processa; ao contrário, mantém com ele uma imbricada relação, sobretudo com a construção dos algoritmos analisados neste trabalho, em razão da relevância dos dados externos e, além disso, na utilização desses algoritmos na execução de atividades sociais.

No que tange ao programador, há duas considerações: a formação dos programadores deve possuir um ensino também de ordem social, que leve em consideração a complexidade social e a responsabilidade de atuação dos programadores. Mais do que ele entender de direito e da ordem jurídica a que está vinculado, é entender que a programação é uma atividade que carrega normatividades que devem ser inclusivas. Outra consideração passa pelo entendimento de que, por meio de um programa, não se criam somente tendências de uso, mas também de normas da comunidade de usuários por meio de seus usos. Observar como essa normatividade pode

20 Nas palavras do porta-voz: "*Our team did test for bias before shipping the model and did not find evidence of racial or gender bias in our testing. But it's clear from these examples that we've got more analysis to do. We'll continue to share what we learn, what actions we take, and will open source our analysis so others can review and replicate*".

ser antecipatória não só de problemas entre usuários ou com outras ordens jurídicas, mas também de problemas de inclusão ou exclusão de usuários, isto é, de problemas discriminatórios. Se, na construção da cidadania, o direito como sistema social deve incluir, o direito como dimensão cultural também não deve escapar disso: o programador e a comunidade de usuários também devem incluir por meio da semântica social do direito.

Dito isso, é possível ensinar tais algoritmos, por meio da regulação dos processos ao longo da programação, a não discriminar, acoplando, em sua programação, os valores incorporados na normatividade jurídica. Essa conclusão é salientada por Kleinberg, Ludwig, Mullainathan e Sustein (2018, p. 114), segundo os quais a proibição de discriminação pode ser implementada pela regulação do processo por meio do qual os algoritmos são construídos.

Ademais, as equipes técnicas, quando conscientes das responsabilidades decorrentes das construções algorítmicas que realizam, podem mitigar a ocorrência da discriminação por meio do reconhecimento de atributos mais capazes de conduzir a tal efeito. Ademais, a criação de times de desenvolvedores mais diversificados pode contribuir para o desenvolvimento de algoritmos mais justos, já que a maioria dos funcionários das grandes empresas de tecnologia são do gênero masculino e brancos (Hao, 2019). Karen Hao aponta que a falta de diversidade nos times de tecnologia se relaciona com os vieses que os algoritmos de IA perpetuam. Esse mesmo problema é apontado no relatório *Discriminating systems: gender, race and power AI*, elaborado em março de 2019 por Sarah M. West, Meredith Whittaker e Kate Crawford. No mesmo sentido, Fei-Fei Li, ex-chefe de IA da Google e professora em Stanford, declara que há muito mais chances de os vieses presentes na sociedade serem transportados para os algoritmos se os times não forem compostos por grupos diversos de engenheiros (Hempel, 2018).

Verificou-se, até aqui, que a discriminação é algo condenável por diversas ordens normativas estatais e até privadas (vide o caso do Twittter exposto acima), mas, a despeito disso, continua a ser um problema persistente. Há grandes problemas em aferir e determinar que uma conduta desse tipo, sobretudo aquelas relacionadas à oferta de emprego ou à demissão por pelo menos duas razões: a primeira, e talvez a mais óbvia, as pessoas costumam mentir acerca das razões que motivaram suas decisões, especialmente quando tais motivações são passíveis de repreensão jurídica e social; em segundo lugar, em certos casos, as pessoas sequer possuem consciência das motivações de seus atos (Kleinberg *et al.*, 2018).

O momento apresenta uma ideia de que os algoritmos já estão enviesados. Todavia, ele pode justamente servir como forma de desenviesar as decisões humanas e promover inclusão. Por exemplo, dada a alta subjetividade na tomada de decisão na contratação de uma pessoa, o algoritmo pode já ser programado para ignorar variáveis que não são preditivas de resultado (sobrenome, cor de pele, sexo, orientação política etc.) (Kleinberg *et al.*, 2018). Assim, toma-se como exemplo uma vendedora que – por ser, em uma sociedade estruturalmente patriarcal, potencialmente mais mal avaliada que um vendedor – pode ser protegida pelo critério em que o algoritmo não leva em conta o seu gênero, o que socialmente já estaria possivelmente no cálculo (Kleinberg *et al.*, 2018). É bem verdade que essas medidas não solucionam integralmente o problema, em razão das "codificações redundantes", apresentadas anteriormente, que conduzem o algoritmo a tomar decisões discriminatórias, a despeito de certas variáveis terem sido eliminadas do *dataset training*. Todavia, diante de um comportamento discriminatório exibido pelo algoritmo – verificado tanto por meio de processos de automonitoramento quanto de supervisão por terceiros –, é possível auditá-lo[21], reestruturá-lo ou simplesmente abandoná-lo e construir novos modelos, buscando compreender como a discriminação foi incutida naquele sistema, o que não seria possível se a decisão fosse exclusivamente humana.

Em um artigo intitulado *An FDA for algorithms*, Tutt (2016, p. 103) afirma que os seres humanos, assim como os algoritmos de aprendizado de máquina, são um exemplo de sistemas imprevisíveis e inexplicáveis. Embora se criem diversas instituições, dentre as quais o direito, para mudar o comportamento humano e adaptá-lo à vida em sociedade, não há como ter plena certeza, isto é, não é possível prever se o comportamento humano realmente seguirá essas instituições e convenções sociais e, quando não, quais os reais motivos que conduziram àquela fuga das normas estabelecidas. Os algoritmos de aprendizado de máquina, por sua vez, embora também se apresentem como sistemas imprevisíveis e inexplicáveis em certas medidas, podem, como argumentam Kleinberg e outros (2018, p. 116), a partir da estruturação de um modelo de observação, identificação e busca de soluções adequado, pautado em determinados princípios técnicos, ser

21 Algumas diretrizes para auditoria desse tipo de sistema são sugeridas pelos autores Pasquale e Citron (2014) para aperfeiçoar a fiscalização do FTC, as quais se referem basicamente à transparência dos sistemas para facilitar a realização de testes, elaboração de relatórios de impacto e análises de risco, dentre outras. De modo similar, as organizações FAT-ML (*Fairness, Accountability and Transparency in Machine Learning Organization*) e ACM (*Association for Computing Machinery*) estabelecem princípios e diretrizes para criação e implementação de algoritmos livres de discriminação.

auditados. Isto é, é possível fazer ao algoritmo perguntas que não podem ser feitas de maneira significativa aos humanos. Em relação às decisões tomadas por algoritmos, é possível aferir, por meio dos mecanismos referidos anteriormente, quais dados foram usados para treinar o modelo algorítmico e quais não foram, além das razões que levaram o sistema àquela decisão de modo que se torna mais fácil para o sistema jurídico descobrir as motivações do sistema (Kleinberg *et al.*, 2018).

Portanto, a criação e implantação de um modelo de observação, identificação e busca de soluções adequado, baseado no diálogo entre os diversos setores da sociedade interessados na regulação dos algoritmos, não "limita simplesmente a possibilidade de discriminação de algoritmos, mas também tem o potencial de transformar os algoritmos em um poderoso contrapeso à discriminação humana e uma força positiva para o bem social" (Kleinberg *et al.*, 2018, p. 115). Talvez, na medida em que os algoritmos de aprendizado de máquina desvelam problemas profundos de discriminação, também sejam eles capazes de mitigar (para não dizer solucionar) esses mesmos problemas, por meio da disponibilização de ferramentas que outrora não existiam e que permitem uma melhor compreensão do fenômeno discriminatório.

CONCLUSÃO

O presente trabalho buscou não somente identificar os problemas de discriminação por meio de algoritmos e sua relação com o princípio da igualdade, mas também verificar respostas que o direito – para além de sua dimensão sistêmica – pode dar para mediar o problema por outras camadas sociais. Por essa razão, é importante entender que a normatividade do direito está também na comunidade de usuários e nas práticas dos programadores, o que revela a importância de impregnar a semântica jurídica da igualdade na seleção e formação de programadores.

A compreensão do direito como prática cultural, impregnada nas diversas relações sociais, além do entendimento do funcionamento dos algoritmos aprendizes nas dinâmicas sociais, possibilita, senão uma resposta para o problema da regulação algorítmica no que concerne às práticas discriminatórias levadas a cabo por esses sistemas, uma via para construção dessa solução.

Os desenvolvedores de sistemas, bem como a comunidade de usuários desses sistemas, organizadas por meio da internet, podem estabelecer

os marcadores adequados para a normatividade da programação, imbuídos de uma semântica pautada e influenciada pelas normas jurídicas voltadas à efetivação de direitos fundamentais, em especial o direito à não discriminação. Isto é, não só o Estado se apresenta como importante pilar da regulação algorítmica, mediante a formalização e institucionalização do direito, mas também, e sobretudo, outros atores diretamente envolvidos na criação e propagação de mecanismos decisórios autônomos, capazes de estruturar a vida em sociedade através da atuação em rede. A compreensão do direito enquanto expressão cultural, nesse sentido, é essencial para que se construam as pontes necessárias entre as normatividades técnicas e jurídicas visando à efetivação de direitos fundamentais.

O trabalho se encerra, mas não evita outras perguntas: que tipo de aprendizados jurídicos os programadores devem ter em sua formação? Considerando que tais atores possuem notória relevância na incorporação de normatividades na programação, em quais outros aspectos específicos e relacionados ao campo jurídico deve ser pautada a formação desses profissionais e como estabelecer essas diretrizes? Além disso, como o Estado pode regular sobre o tema sem retirar a dinâmica de inovação de atores privados? Como a autorregulação privada de plataformas de internet devem se ocupar sobre o tema? As respostas para essas perguntas exigem outras agendas de pesquisa e novas investigações, razão pela qual este trabalho deve ser visto não como o ponto final que encerra a discussão, mas como um meio que, ao mesmo tempo em que apresenta (possibilidades de) soluções para determinado problema, revela e direciona novas perguntas sobre outros aspectos relevantes.

REFERÊNCIAS

ALI, Muhammad; SAPIEZYNSKI, Piotr; BOGEN, Miranda; KOROLOVA, Aleksandra; MISLOVE, Alan; RIEKE, Aaron. *Discrimination through optimization. Proceedings of the ACM on human-computer interaction*. [s.l.], v. 3, n. , p. 1-30, 7 nov. 2019. Association for Computing Machinery (ACM). Disponível em: http://dx.doi.org/10.1145/3359301.

ALPAYDIN, Ethem. *Introduction to machine learning*. 3. ed. Massachusetts: MIT Press, 2014.

ARTOSI, Alberto. Technical normativity. In: *Italian Philosophy of Technology*. Springer, Cham, p. 149-160, 2021.

ASSMANN, Jan. *Das kulturelle Gedächtnis*: Schrift, Erinnerung und politische Identität in frühen Hochkulturen. München: Beck, 1992.

BAROCAS, Solon; SELBST, Andrew D. Big data's disparate impact. *Calif. L. Rev.*, v. 104, p. 671, 2016.

BUOLAMWINI, Joy. How I'm fighting bias in algorithm. *TEDx BeaconStreet*, 2016, TED Talks. Disponível em: https://www.ted.com/talks/joy_buolamwini_how_i_m_fighting_bias_in_algorithms. Acesso em: 10 set. 2020.

CALDERS, Toon; ŽLIOBAITĖ, Indrė. Why unbiased computational processes can lead to discriminative decision procedures. In: CUSTERS, Bart et al (Ed.). *Discrimination and privacy in the information society*: data mining and profiling in large databases. Berlin: Springer-Verlag, p. 3-26, 2013. (Studies in Applied Philosophy, Epistemology and Rational Ethics 3).

CALO, Ryan. *Artificial intelligence policy*: a primer and roadmap. 8 ago. 2017. Disponível em: https://ssrn.com/abstract=3015350. Acesso em: 17 jul. 2020.

CITRON, Danielle Keats; PASQUALE, Frank A. The scored society: due process for automated predictions. *Washington Law Review*, Vol. 89, 2014, p. 1-, U of Maryland Legal Studies Research Paper No. 2014-8, 2014.

CORTIZ, Diogo. *Curso de inteligência artificial para todos (Parte 1)*. Diogo Cortiz, 2020. YouTube. Disponível em: https://youtu.be/Ze-Q6ZNWpco. Acesso em: 24 mar. 2020.

CUSTERS, Bart. Data dilemmas in the information society: introduction and overview. In: CUSTERS, Bart et al (Ed.). *Discrimination and privacy in the information society*: data mining and profiling in large databases. Berlin: Springer-Verlag, p. 3-26, 2013. (Studies in Applied Philosophy, Epistemology and Rational Ethics 3).

DIAKOPOULOS, Nicholas et al. *Principles for accountable algorithms and a social impact statement for algorithms*. Fairness, accountability, and transparency in machine learning. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 1º out. 2020.

DOMINGOS, Pedro. *O algoritmo mestre*. São Paulo: Novatec, 2017.

ERNST, Christian. *Algorithmische Entscheidungsfindung und personenbezogene Daten Privatdozent*, Juristen Zeitung 72(21), 2017.

ESPOSITO, Elena. *Artificial communication?* The production of contingency by algorithms. In: 46, 4 Zeitschrift für Soziologie, 2017, p. 249-265.

FLORIDI, Luciano. *The Onlife Manifesto*: being human in a hyperconnected era. New York: Springer Nature, 2015.

GEERTZ, Clifford. *Local knowledge*: further essays in interpretative anthropology. New York: Basic Books, 1983.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. *Data mining*: um guia prático. Rio de Janeiro: Elsevier, 2005.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep learning*. MIT Press. 2016. Disponível em: http://www.deeplearningbook.org. Acesso em: 18 ago. 2020.

HOFFMANN-RIEM, Wolfgang. Verhaltenssteuerung durch Algorithmen – Eine Herausforderung für das Recht. In: *Archiv des oeffentlichen Rechts*, Volume 142, Number 1, January 2017Mohr Siebeck: Tübingen, 2017.

_____. Inteligência artificial como oportunidade para a regulação jurídica. *Direito Público*, [s.l.], v. 16, n. 90, dez. 2019. ISSN 2236-1766. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3756. Acesso em: 11 ago. 2020.

JUST, Natascha. Governing online platforms: competition policy in times of platformization. In: *Telecommunications Policy, Telecommunications Policy*, Volume 42, Issue 5, June 2018, Pages 386-394. Disponível em: https://doi.org/10.1016/j.telpol.2018.02.006.

_____; LATZER, Michael. Governance by algorithms: reality construction by algorithmic selection on the internet. Accepted manuscript forthcoming in Media. *Culture & Society*, 2016, s. 2. Disponível em: http://mediachange.ch/media/pdf/publications/Just_Latzer2016_Governance_by_Algorithms_Reality_Construction.pdf.

KAPLAN, Andreas; HAENLEIN, Michael. Siri, Siri, in my hand: who's the fairest in the land? On interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, [s.l.], v. 62, n. 1, p. 15-25, jan. 2019. Elsevier BV. http://dx.doi.org/10.1016/j.bushor.2018.08.004.

LADEUR, Karl-Heinz. Lernfähigkeit des Rechts und Lernfähigkeit durch Recht, Erwiderung auf J. Nocke. In: GÖRLITZ, Axel; VOIGT, Rüdiger (Hrsg.). *Postinterventionistisches Recht*, Pfaffenweiler: Centaurus 1990 (Jahresschrift für Rechtspolitologie, Bd. 4), p. 141 ss.

_____. *Die Textualität des Rechts*: Zur poststrukturalistischen Kritik des Rechts. Velbrück Wissenschaft. Weilerwist, 2016.

_____. *Recht – Wissen – Kultur*. Die fragmentierte Ordnung. Berlin Duncker & Humblot, 2016.

LERMAN, Jonas. Big data and its exclusions. *Stan. L. Rev. Online*, v. 66, p. 55, 2013. Disponível em: https://review.law.stanford.edu/wp-content/uploads/sites/3/2016/08/66_stanlrevonline_55_lerman.pdf. Acesso em: 10 jul. 2020.

LESSIG, Lawrence. *Code*: Version 2.0, New York: Basic Books, 2006.

MAINI, Vishal; SABRI, Samer. *Machine learning for humans*. Disponível em: https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf. Acesso em: 17 ago. 2020.

MARR, Bernard. How much data do we create every day? The mind-blowing stats everyone should read. *Forbes*, 18 maio 2018. Disponível em: https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#7574ae7d60ba. Acesso em: 1º abr. 2020.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data*: a revolution that will transform how we live, work and think. New York: 2013.

MENDES, Laura Schertel. *Privacidade, proteção de dados e defesa do consumidor*: linhas gerais de um novo direito fundamental. São Paulo: Saraiva, 2014.

MITCHELL, Tom M. *Machine learning*. McGraw-Hill Science, 1997.

NEGÓCIO, Ramon de Vasconcelos. *Vom Fremddruck zur Selbstbeschränkung*: Das Problem der Verarbeitung juridischer Normativität durch Internet-Intermediäre. Baden-Baden: Nomos, 2020.

NEVES, Marcelo. Entre subintegração e sobreintegração: a cidadania inexistente. *Dados – Revista de Ciências Sociais*, Rio de Janeiro, v. 37, n. 2, p. 253-275, 1994.

_____. *Transconstitucionalismo*. São Paulo: WMF Martins Fontes, 2009.

O'NEIL, Cathy. *Weapons of math destruction*: how big data increases inequality and threatens democracy. New York: Crown Publishers, 2016.

PAGER, Devah; SHEPHERD, Hana. The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annual Review Of Sociology*, [s.l.], v. 34, n. 1, p. 181-209, ago. 2008. Annual Reviews. Disponível em: http://dx.doi.org/10.1146/annurev.soc.33.040406.131740.

PEIXOTO, Fabiano Hartmann; SILVA, Roberta Zumblick Martins da. *Inteligência artificial e Direito*. Curitiba: Alteridade, 2019.

PIOVESAN, Flávia. Ações afirmativas no Brasil: desafios e perspectivas. *Revista Estudos Feministas*, Florianópolis, 16(3): 424, set./dez. 2008.

RAWLS, John. *Uma teoria da Justiça*. Trad. Almiro Pisetta e Lenita M. R. Esteves. São Paulo: Martins Fontes, 2008.

ROSEN, Lawrence. *Law as culture*: an invitation. New Jersey: Princeton University Press, 2006.

SCHULZ, W.; DANKERT, K. *Die Macht der Informationsintermediäre* – Erscheinungsformen, Strukturen, Regulierungsoptionen. Bonn: Friedrich-Ebert-Stiftung, 2016.

SCOTT, Joan W. O enigma da igualdade. *Revista Estudos Feministas*, [s.l.], v. 13, n. 1, p. 11-30, abr. 2005. FapUNIFESP (SciELO). Disponível em: http://dx.doi.org/10.1590/s0104-026x2005000100002.

TISCHBIREK, Alexander. Artificial intelligence and discrimination: discriminating against discriminatory systems. In: *Regulating Artificial Intelligence*, ed. Thomas Wischmeyer e Timo Rademacher, 2019.

TURING, Alan. Computing machinery and intelligence. *Mind*, New Series, v. 59, n. 236, p. 433-460, out. 1950.

TUTT, Andrew. An FDA for algorithms. *Administrative Law Review*, 83 (2017). Disponível em: https://ssrn.com/abstract=2747994. Acesso em: 7 nov. 2020.

VESTING, Thomas. The autonomy of law and the formation of network standards. *German Law Journal*, Vol. 05, No. 06, p. 639-668, 2004.

_____. *Die Medien des Rechts*: Computernetzwerke. Weilerswist: Velbrück Wissenschaft, 2015.

WEST, S. M.; WHITTAKER, M.; CRAWFORD, K. *Discriminating Systems*: Gender, Race and Power in AI. AI Now Institute, 2019. Disponível em: https://ainowinstitute.org/discriminatingsystems.html. Acesso em: 1º nov. 2020.

Sobre os autores:

Alan Duarte | *E-mail:* duarttacademic@gmail.com

Mestrando em Direito Constitucional pela Universidade Federal do Ceará. Pós-Graduado em Direito, Tecnologia e Inovação (Instituto New Law). Advogado. Graduado em Direito pelo Centro Universitário 7 de Setembro (2020).

Ramon de Vasconcelos Negócio | *E-mail:* negocioramon@gmail.com

Doutor em Teoria do Direito (2019) pela Goethe-Universität (Frankfurt am Main). Mestre em Direito do Estado, na subárea de Direito Constitucional (2011), pela Pontifícia Universidade Católica de São Paulo. Graduado em Direito pela Universidade de Fortaleza (2007).

Data de submissão: 14 de agosto de 2021.
Data de aceite: 10 de janeiro de 2022.

Discriminação Algorítmica e Inclusão em Sistemas de Inteligência Artificial – Uma Reflexão sob a Ótica dos Direitos da Criança no Ambiente Digital

Algorithmic Discrimination and Inclusion in Artificial Intelligence Systems – A Reflection from the Perspective of Children's Rights in the Digital Environment

ISABELLA VIEIRA MACHADO HENRIQUES[1]
Pontifícia Universidade Católica de São Paulo (PUC/SP).

INÊS VITORINO SAMPAIO[2]
Universidade Federal do Ceará (UFC).

RESUMO: O artigo trata de explorar e sistematizar o tema da discriminação algorítmica e da inclusão em sistemas de inteligência artificial, abordando-o sob a ótica dos direitos das crianças no ambiente digital no Brasil. Por meio de revisão bibliográfica em articulação com a análise factual de casos notórios de discriminação por sistemas de inteligência artificial, contextualiza a discriminação algorítmica de maneira ampla e em relação a crianças. Com base em revisão bibliográfica e análise documental de legislações como a Constituição Federal e a Lei Geral de Proteção de Dados Pessoais, além de cartas internacionais, aborda os direitos fundamentais, especialmente o direito à inclusão e à não discriminação e a sua incidência na infância, bem como a relação entre os princípios éticos associados à inteligência artificial e os princípios e conceitos da proteção de dados pessoais, em termos genéricos e no recorte específico dos direitos das crianças no ambiente digital. Por fim, aponta possíveis caminhos para a solução em relação à discriminação algorítmica que acomete pessoas adultas, mas impacta sobremaneira crianças, no sentido de que as múltiplas infâncias sejam incluídas no ciberespaço com a garantia de seus direitos humanos.

PALAVRAS-CHAVE: Discriminação; criança; inclusão; inteligência artificial.

1 Orcid: http://orcid.org/0000-0003-1911-9223.
2 Orcid: http://orcid.org/0000-0001-7507-4053.

ABSTRACT: The article explores and systematizes the theme of algorithmic discrimination and the inclusion in Artificial Intelligence systems, approaching it from the perspective of children's rights in the digital environment in Brazil. Through a literature review in conjunction with the factual analysis of notorious cases of discrimination by Artificial Intelligence systems, it contextualizes algorithmic discrimination in a broad manner and in relation to children. Based on a literature review and document analysis of legislation such as the Federal Constitution and the General Law for the Protection of Personal Data, in addition to international charters, it addresses fundamental rights, especially the right to inclusion and non-discrimination and its incidence in childhood, as well as the relationship between the ethical principles associated with Artificial Intelligence and the principles and concepts of personal data protection, in generic terms and in the specific scope of children's rights in the digital environment. Finally, it points out possible paths for a solution in relation to the algorithmic discrimination that affects adults, but which has a major impact on children, in the sense that multiple childhoods are included in cyberspace with the guarantee of their human rights.

KEYWORDS: Discrimination; child; inclusion; artificial intelligence.

INTRODUÇÃO

Nos últimos anos, o mundo mudou e tem se transformado em uma velocidade sem precedentes. Grande parte das mudanças vistas a olhos nus tem relação com algo que não se enxerga facilmente, mas é onipresente e onisciente: os sistemas de inteligência artificial ("IA") (Harari, 2016). Nos dias atuais, as inovações tecnológicas advindas desses sistemas estão presentes na vida cotidiana de grande parte das pessoas ao redor do planeta, como, por exemplo, em redes sociais, sistemas de busca na internet, *chatbots*, recomendações de filmes nos serviços de *streaming*, assistentes virtuais, brinquedos inteligentes, *wearables*[3], sistemas de reconhecimento facial, aplicações na agricultura e na indústria, robôs e plataformas educacionais, exames médicos e até em carros autônomos.

Essa expansão e, consequente, popularização do uso de sistemas de IA, que concedem a uma entidade não natural habilidades para a tomada de decisões por meio de um processo avaliativo (Turner, 2019), instaura-se acompanhada de imensos desafios para toda a sociedade, governos e

3 Dispositivos eletrônicos vestíveis como, por exemplo, relógios conectados que monitoram a saúde da pessoa usuária. Disponível em: https://www.softeq.com/blog/5-ways-ai-powered-wearable-devices-are-rocking-the-healthcare-industry. Acesso em: 12 set. 2021.

empresas. Estes desafios envolvem diversos campos, como saúde, educação, segurança, trabalho, justiça e transporte, entre outros. Dizem respeito a questões éticas, filosóficas, regulatórias e tecnológicas, que demandam uma compreensão integrada dos fenômenos que interferem nos comportamentos humanos (Souza; Oliveira, 2019).

No princípio, acreditava-se que a IA abriria caminhos nunca antes explorados pelos seres humanos. Hoje, contudo, existem versões sofisticadas que, simplesmente, exploram os caminhos já trilhados, mas com muitíssimo mais eficiência (Getschko, 2021). Nesse percurso, parcela importante dessa inovação tecnológica tem sido associada a uma busca por identificar e promover benefícios da IA para a humanidade, em especial, relacionados à melhoria da condição de vida humana.

Por outro lado, é certo que a inteligência artificial apresenta riscos variados, sendo que um dos maiores, já amplamente reconhecido, diz respeito à delegação de decisões para a máquina (Mulholland; Frajhof, 2019). Diferente da utilização da IA como ferramenta, que auxilia o ser humano na tomada de decisões – em relação ao que há um consenso sobre a sua pertinência e adequação –, a delegação da tomada de decisão pela máquina pode trazer graves consequências e envolve uma série de desafios, dentre os quais, potenciais violações a direitos fundamentais, como por meio da existência de vieses nas resultantes dessas decisões (Doneda; Mendes; Souza; Andrade, 2018).

Não faltam exemplos de implicações discriminatórias de decisões algorítmicas. A linha do tempo do racismo algorítmico não deixa dúvidas a esse respeito: atualizada até 2021, apresenta casos desde 2010 (Silva, 2021). O recente documentário Coded Bias, que tem como protagonista a cientista de computação do Instituto de Tecnologia de Massachusetts (MIT) Joy Buolamwini, cofundadora do movimento Algorithmic Justice League[4], reforça a atualidade e a magnitude do problema.

Em razão dos potenciais danos envolvidos e dos inúmeros casos de discriminação vindos a público, a discriminação algorítmica associada a sistemas de IA, em especial, de grupos historicamente discriminados, tem sido motivo de grande preocupação de indivíduos e grupos sociais, como a comunidade internacional, os organismos multilaterais e as mais variadas

4 Disponível em: https://www.ajl.org/. Acesso em: 6 set. 2021.

instituições, públicas e privadas, consumidoras e desenvolvedoras de tecnologias.

A propósito, o relatório da Relatora Especial, Tendayi Achiume, para as formas contemporâneas de racismo, discriminação racial, xenofobia e intolerância relativa, intitulado *Racial Discrimination and Emerging Digital Technologies: a Human Right Analysis*, apresenta um robusto diagnóstico combinado com recomendações e obrigações aos Estados-Partes, para que proíbam, combatam e previnam a discriminação no desenvolvimento e uso das novas tecnologias digitais (ONU, 2020). Iniciativas de empresas privadas, como o projeto Crowdsource do Google[5], que disponibiliza *site* e aplicativo para que as pessoas adicionem conteúdos regionais visando a expandir sua base de dados, em busca de maior inclusão, têm sido mais frequentemente apresentadas.

É nesse contexto mais amplo de preocupações sobre os usos de sistemas de IA que ganham relevo a questão dos direitos digitais e a consideração de aspectos éticos relacionados ao melhor interesse das crianças. No tocante a esse enfoque, entre as várias ações dos organismos multilaterais, destaca-se o *Policy Guidance on AI for Children*, que reforça a necessidade de se conceber uma IA para todas as crianças, que não as discrimine, mas, ao contrário, priorize e se esteie no princípio da justiça, garantindo a inclusão *de* e *para* crianças (Unicef, 2020).

A abordagem dessa questão, evidentemente, não pode se reduzir a mera retirada de viés, mas deve se pautar pelo propósito de se inserir um viés humanista nas decisões das máquinas (Getschko, 2021). Ainda assim, não se deve minimizar o fato de que resultantes discriminatórias têm sido constantes, evidenciando a urgência pelo enfrentamento dessa questão, especialmente em se tratando do público infantil, que é mais vulnerável e vivenciará por maior tempo as implicações dos sistemas de IA. Urge, portanto, tomar medidas concretas no combate à discriminação e na direção de maior inclusão das multiplicidades étnicas, raciais, etárias, nacionais e de gênero no ambiente digital e na tomada de decisão algorítmica, inclusive em relação às crianças. Daí a necessidade de o ciberespaço ser constrangido por leis, normas sociais, mercado e códigos de *software* da própria tecnologia, a fim de garantir valores que ressoem a tradição humanística contemporânea e vislumbrem um regime mais liberal e menos controlador (Lessig, 1999).

5 Disponível em: https://crowdsource.google.com/. Acesso em: 7 set. 2021.

Frente ao exposto, o objetivo do presente artigo é explorar e sistematizar o tema da discriminação algorítmica e da inclusão em sistemas de IA, abordando-o sob a ótica dos direitos das crianças[6] no ambiente digital no Brasil. Também pretende apontar caminhos que contribuam para enfrentar essa questão.

O artigo está dividido em três partes, além desta introdução e das considerações finais. Na primeira parte, além da revisão bibliográfica em articulação com a análise factual de casos notórios de discriminação por sistemas de IA, contextualiza-se a discriminação algorítmica de maneira ampla e em relação a crianças.

Na segunda parte, com base na revisão bibliográfica e na análise documental de legislações como a Constituição Federal e a Lei Geral de Proteção de Dados Pessoais – Lei nº 13.709/2018 – ("LGPD"), além de cartas internacionais, são abordados os direitos fundamentais das crianças, especialmente o direito à inclusão e à não discriminação, e a relação entre os princípios éticos associados à IA e os princípios e conceitos da proteção de dados pessoais, em termos genéricos e no recorte específico dos direitos das crianças no ambiente digital.

Na terceira e última parte, são apontados possíveis caminhos para a solução em relação à discriminação algorítmica que acomete pessoas adultas, mas que impacta sobremaneira crianças, no sentido de que as múltiplas infâncias sejam incluídas no ciberespaço com a garantia de seus direitos humanos.

1 CONTEXTUALIZAÇÃO DA DISCRIMINAÇÃO ALGORÍTMICA EM SISTEMAS DE IA E COM RECORTE NAS INFÂNCIAS

A discriminação algorítmica de que trata este artigo é originada em sistemas de IA das mais diversas e variadas características quanto ao seu uso e propósito. No entanto, têm em comum a utilização de recursos que possibilitam a predição, pois permitem que *softwares* aprendam determinados padrões extraídos dos dados que os alimentam, por meio da combinação de grandes quantidades de dados com algoritmos inteligentes. Tais algoritmos

6 Este artigo vale-se do conceito de "criança" previsto na Convenção sobre os Direitos da Criança da Organização das Nações Unidas (ONU), que foi recepcionada no Brasil pelo Decreto nº 99.710/1990, no sentido de englobar pessoas de 0 a 18 anos. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/1990-1994/d99710.htm. Acesso em: 6 set. 2021.

consistem em etapas nas quais são completadas tarefas descritas de maneira precisa o bastante para um computador realizá-las (Cormen, 2013), ou seja, que são conjuntos de instruções para realizar tarefas que produzem resultados finais tendo por base algum ponto de partida (Doneda; Almeida, 2016).

É por isso que a IA pode ser considerada um subcampo da ciência da computação, focada na construção de máquinas e *softwares* que podem, de alguma forma, mimetizar comportamentos considerados inteligentes (SAS Institute, 2021). Vale dizer que não existe uma única definição que seja indistintamente aceita para a IA, mas o termo tem sido usado quando uma máquina ou sistema executa tarefas que normalmente exigiriam capacidade cerebral humana ou biológica para realizar, como compreender linguagem falada, aprender comportamentos ou resolver problemas. Atualmente, há uma grande variedade de sistemas de IA, os quais, de modo geral, consistem em computadores que executam algoritmos baseados em dados (The Alan Turing Institute, 2021).

Sistemas de IA podem interagir com as pessoas e atuar nos seus ambientes direta ou indiretamente, assim como operar de forma autônoma e adaptar seu comportamento aprendendo sobre o contexto. Em outras palavras, são sistemas baseados em máquina que, dado um conjunto de objetivos definidos pelo ser humano, têm a capacidade de fazer previsões e recomendações ou de tomar decisões que influenciam ambientes reais ou virtuais (Unicef, 2020).

Importa, ainda, mencionar que a capacidade aumentada de processamento dos dados alcançada via *machine learning* destravou um ponto crucial – a velocidade para a tomada de decisão – considerada um gargalo para qualquer tipo de automação. Assim, se os algoritmos são considerados o motor, certamente os dados são o combustível da atual revolução tecnológica (Bigonha, 2018).

Por isso que se diz que um algoritmo é tão bom quanto os dados que o alimentam e que seu uso apresenta riscos não evidentes, derivados especialmente dos seguintes fatores: a possibilidade de promoverem discriminação, ainda que sem intenção, o que acontece quando as bases de dados utilizadas para o treinamento remetem a vieses presentes na sociedade; o emprego de base de dados incompletas, e a opacidade na sua forma de atuação, consequência de determinadas técnicas de aprendizado de máquina (Ferrari, 2020).

Em relação aos referidos riscos, no que diz respeito ao tema do presente artigo, consoante previsto na Declaração de Toronto (Anistia Internacional; Access Now, 2018), a discriminação[7] pode ser conceituada como:

> Qualquer distinção, exclusão, restrição ou preferência baseada em qualquer fundamento, como raça, cor, sexo, idioma, religião, opinião política ou outra, origem nacional ou social, propriedade, nascimento ou outra condição de nascimento, e que tem por objetivo ou efeito anular ou impedir o reconhecimento, gozo ou exercício por todas as pessoas, em igualdade de condições, de todos os direitos e liberdades.[8]

Resultados discriminatórios podem, ainda, decorrer do fato de alguém "pertencer a determinado grupo e ser julgado a partir das características desse grupo; um cenário no qual as características individuais de uma pessoa são desconsideradas, e aquela pessoa é vista somente como um membro de um dado conjunto de pessoas" (Mendes; Mattiuzzo, 2019, p. 47). Nesse sentido, uma afirmação pode ser descrita como preconceituosa quando se baseia em generalizações estatísticas inconsistentes, mas também quando se refere a generalizações estatisticamente consistentes, mas não universais, na medida em que as pessoas merecem ser tratadas como indivíduos, e não apenas como membros de um grupo (Mendes; Mattiuzzo, 2019).

No campo da IA, algoritmos podem possuir vieses discriminatórios embutidos quando os vieses dos seus desenvolvedores forem passados à máquina, ainda que sem maiores percepções ou por má intenção deliberada, senão, por erro humano de programação. Contudo, ainda que não possuam vieses discriminatórios embutidos, algoritmos têm a capacidade de gerar resultados insatisfatórios e flagrantemente preconceituosos, se forem alimentados por dados com conceitos e valores repletos de vieses, passando a com eles aprender.

Daí a noção de que a máquina, por si só, não é preconceituosa, mas sim os seres humanos que a desenvolvem e a alimentam, ainda que, como tecnologias de classificação que diferenciam, classificam e categorizam, sistemas de IA sejam discriminatórios por natureza (West; Whittaker; Crawford, 2019).

---

7 Trata-se de uma lista não exaustiva, pois o Alto Comissariado das Nações Unidas para os Direitos Humanos já reconheceu a necessidade de prevenir discriminação contra classes adicionais.

8 Tradução livre do item 13 da Declaração.

As Professoras Laura Mendes e Marcela Mattiuzzo (2019), a propósito, apresentam a seguinte tipologia das discriminações algorítmicas: (i) por erro estatístico; (ii) por generalização incorreta; (iii) por uso de informações sensíveis; e (iv) por limitação do exercício de direitos (Mendes; Mattiuzzo, 2019).

As tecnologias digitais e a própria IA não são neutras, refletem valores e interesses de quem influencia a sua concepção e seu uso, bem como são fruto das mesmas estruturas de poder e desigualdade que operam na sociedade. Não só têm a capacidade de reproduzir, mas exacerbar as desigualdades existentes em vários contextos, até porque a automatização de discriminações históricas pode acarretar danos de alcance massivo. Sob o ponto de vista das crianças, podem reverberar exclusões e discriminações em uma fase que é de aprendizagem de suas leituras de mundo. Além de implicarem situações de sofrimento, projetam-se num escopo de tempo mais vasto (da infância à vida adulta), dificultando ainda mais a resolução desses problemas.

São inúmeras as circunstâncias que geram discriminação, sendo muito recorrentes: as raciais, de gênero, etárias e socioeconômicas. Especialmente, quando se dá a interseccionalidade, com a sobreposição dessas identidades sociais em situações de opressão e dominação (ONU, 2020). A respeito da discriminação na faixa etária das crianças, notoriamente mais vulnerável, por serem pessoas que vivenciam um período peculiar de desenvolvimento biopsicossocial (Piaget; Inhelder, 2021), o Comentário Geral nº 25 sobre os direitos das crianças em relação ao ambiente digital[9], elaborado pelo Comitê dos Direitos da Criança da ONU (ONU, 2021), apresenta as seguintes situações prejudiciais nas quais crianças podem ser discriminadas:

> Crianças podem ser discriminadas por serem excluídas do uso de tecnologias e serviços digitais ou por receberem comunicações de ódio ou tratamento injusto no uso dessas tecnologias. Outras formas de discriminação podem surgir quando processos automatizados que resultem em filtragem de informações, perfilamento ou tomada de decisões são baseados em dados tendenciosos, parciais ou obtidos de forma injusta em relação a uma criança.[10]

9 O mesmo Comentário Geral nº 25 assim define ambiente digital: "Tecnologias da informação e comunicação, incluindo redes, conteúdos, serviços e aplicativos digitais, dispositivos e ambientes conectados; realidade virtual e aumentada; robótica; *inteligência artificial*; sistemas automatizados, algoritmos e análise de dados; biometria e tecnologia de implantes" (ONU, 2021 – grifos nossos).

10 Tradução livre do item 10 do Comentário Geral nº 25.

Crianças representam, no mundo, um terço dos usuários de internet, sendo que jovens entre 15 e 24 anos representam a faixa etária mais conectada: 71% estão *online*, em comparação com 48% da população total (Unicef, 2017). Isso faz crianças, adolescentes e jovens estarem, cada vez mais, em contato com aplicações de IA. Ainda que 29% de jovens entre 15 e 25 anos (346 milhões) não tenham acesso à internet[11], sendo os jovens africanos os menos conectados do mundo: 60% deles não estão *online*, comparados a somente 4% na Europa (ITU, 2020).

No Brasil, 89% das crianças de 9 a 17 anos são usuárias de internet[12], e o telefone celular segue predominante como dispositivo de acesso à rede para 95%. São 3 milhões de crianças não usuárias, sendo que 1,4 milhões nunca acessaram a internet (Cetic.br, 2019). Há, pois, uma questão de discriminação via exclusão digital e social que é anterior a qualquer processo de discriminação por IA, mas que pode, como já salientado, ser intensificado via sistemas de IA que não tragam, em seu desenho, o compromisso ético com o melhor interesse da criança.

O país possui 69,8 milhões de pessoas de 0-19 anos, que representam 33% da população. 46,8% de crianças entre 0 e 14 anos vivem em condição domiciliar de baixa renda. Enquanto a região Norte possui a maior proporção de crianças no país, acima de 40%, mais de 20% dos seus estabelecimentos de educação básica declaram não possuir acesso ao esgoto sanitário (Fundação Abrinq, 2021). Em razão da desigualdade estrutural, uma criança demora até 9 gerações para deixar a faixa dos 10% mais pobres e chegar à renda média do país (OECD, 2018).

São múltiplas as infâncias no Brasil, em especial considerados fatores socioeconômicos e culturais. Com variadas culturas locais e regionais, o português como língua oficial, possui, ainda, a língua brasileira de sinais (Libras), além de línguas indígenas e dialetos regionais. São infâncias que possuem diferenças socioeconômicas e culturais, além de diferenças de gênero e etárias. Englobam crianças com e sem deficiências; negras, brancas ou amarelas; indígenas, quilombolas, ribeirinhas, refugiadas, entre outras tantas origens; de diferentes regiões e mesmo nacionalidades, urbanas, li-

11 3,7 bilhões de pessoas no mundo não têm acesso à internet, sendo que, em países pobres, 17% da população rural vive em áreas sem cobertura de internet e 19% possui cobertura apenas por uma rede de 2G (ITU, 2020).

12 83% assistiu a vídeos, programas, filmes ou séries. 76% pesquisou na internet para fazer trabalhos escolares. 68% usou redes sociais. 59% baixou músicas e filmes. 57% jogou *online* conectado com outros jogadores. Nas camadas mais pobres, 73% acessam a internet exclusivamente pelo celular (CETIC, 2019).

torâneas ou rurais. São distintas umas das outras e iguais no ser criança. São similares na vivência das fases de desenvolvimento biopsicossocial, bem como nas suas respectivas e inerentes características físicas e cognitivas (Marino; Chicaro, 2019) tão peculiares que lhes garantem um direito supranacional[13] à proteção especial. Ao mesmo tempo, são diferentes porque vivenciam jornadas distintas, individual e coletivamente, relacionadas ao ambiente no qual vivem, àquilo a que têm acesso, à forma como conhecem e relacionam-se com o mundo e como nele conseguem interagir (Gardner, 2018).

Como em outros países do sul global, as infâncias no Brasil são atravessadas pela desigualdade social e, também por isso, encaram mais barreiras e riscos – inclusive da discriminação – para usufruir o ambiente digital na sua maior potência e conforme o seu melhor interesse (ECPAT International, 2020). Até porque esse ambiente digital, no qual há a prevalência de um modelo de negócio baseado em uma "vigilância líquida" (Bauman; Lyon, 2014) ou no "capitalismo de vigilância" (Zuboff, 2019), está inserido em um ambiente mais amplo que é de uma sociedade individualista, da informação e do espetáculo ou da hipermodernidade, do hiperconsumo e hiperconexão, naquela que Gilles Lipovetsky chama de "a era do vazio" (Lipovetsky, 2005).

É nesse contexto que exemplos de discriminação resultantes de decisões tomadas por máquinas pulverizam-se. O relatório de pesquisa do National Institute of Standards and Technology (NIST), que revisou 189 algoritmos de reconhecimento facial de 99 desenvolvedores em todo o mundo, apontou que muitos destes algoritmos eram de 10 a 100 vezes mais propensos a identificar imprecisamente uma fotografia de um rosto negro ou asiático, em comparação com um branco (Grother; Ngan; Hanaoka, 2019). Da mesma forma, o notório caso do Compas – Correctional Offender Management Profiling for Alternative Sanctions[14] –, após reportagem da *ProPublica*, agência de jornalismo investigativo, teve seu viés discriminatório e racista alardeado (Angwin; Larson; Mattu; Kirchner, 2016).

Outro caso igualmente conhecido é o do *chatbot* Tay, desenvolvido pela Microsoft, que, em pouco tempo, adquiriu uma personalidade extremamente agressiva e preconceituosa, tornando-se uma espécie de nazista

13 A Convenção sobre os Direitos da Criança da ONU.

14 *Software* privado usado para auxiliar a dosimetria das penas estipuladas pelo Judiciário, nos Estados Unidos.

virtual, por ter tido seu sistema de IA manipulado por internautas[15]. Com vieses problemáticos semelhantes, a ferramenta de recrutamento da Amazon com IA que discriminava candidatas mulheres[16] e o *AppleCard* tornaram-se alvo de investigação pelo Departamento de Serviços Financeiros de Nova Iorque por usar algoritmo sexista[17].

Também acusados de discriminação o sistema de tradução do Google, em relação ao gênero de palavras em idiomas que possuem o gênero neutro, como "*doctor*", traduzido no português para "o médico" e "*nurse*" para "a enfermeira"[18], e o sistema de definição de palavras da mesma empresa, que designava a palavra "professora" como "prostituta com quem adolescentes se iniciam na vida sexual"[19].

Ainda que esses casos não se refiram, sobretudo, às crianças, impactam diretamente nelas, pois têm incidência sobre adultos integrantes do seu círculo familiar, de pertença racial, étnica, socioeconômica etc. Ademais, no caso específico do sistema de tradução supracitado, a incidência no processo de formação da criança que realiza suas pesquisas é notória. Isto sem falar nos sistemas de busca que ela acessa. Ao fazê-lo, a criança é monitorada em suas práticas cotidianas de descoberta do mundo, tendo seus dados colhidos para usos diversos, sobre os quais têm pouco ou nenhum controle. Nesse processo, ela acessa conteúdos frequentemente impulsionados por lógicas comerciais, nem sempre atentas ao seu melhor interesse, o que pode reverberar no seu acesso a textos e imagens prejudiciais à sua formação, a exemplo de discursos de ódio, sexistas etc.

Em relação a crianças, de modo específico, inúmeras situações de discriminação algorítmica têm sido também verificadas, como, por exemplo, no caso do sistema de reconhecimento facial de Buenos Aires, com foco na segurança pública, cujo suposto infrator mais jovem identificado, citado por crimes de ferimentos graves contra pessoas, teria menos de quatro anos! Segundo o MIT Technology Review, em testes anteriores realizados

15 Disponível em: https://www.tecmundo.com.br/inteligencia-artificial/102835-microsoft-explica-episodio-chatbot-racista-diz-tay-deve-voltar.htm. Acesso em: 18 set. 2021.

16 Disponível em: https://tecnoblog.net/meiobit/391571/ferramenta-de-recrutamento-amazon-ai-discriminava-mulheres/. Acesso em: 18 set. 2021.

17 Disponível em: https://www.tecmundo.com.br/mercado/147626-apple-card-alvo-investigacao-usar-algoritmo-sexista.htm. Acesso em: 18 set. 2021.

18 Disponível em: https://www.tecmundo.com.br/internet/136939-google-quer-combater-estereotipos-genero-traducoes-google-tradutor.htm. Acesso em: 18 set. 2021.

19 Disponível em: https://olhardigital.com.br/2019/10/23/noticias/google-remove-definicao-de-professora-como-prostituta-no-dicionario/. Acesso em: 18 set. 2021.

pelo governo dos Estados Unidos, o algoritmo utilizado neste sistema teria um desempenho pior, por um fator de seis, em crianças com idades entre 10 e 16 em relação a adultos de 24 a 40 anos. Isso provavelmente porque, de acordo com documentos oficiais, o sistema teria sido testado apenas em rostos adultos de funcionários do governo municipal de Buenos Aires antes da sua aquisição[20].

Sistemas de reconhecimento facial, mesmo sob as condições ideais de laboratório, são considerados ruins para lidar com crianças, justamente porque são treinados e testados, na sua imensa maioria, em adultos. Ademais, a maior parte das ferramentas utilizadas hoje no sul global foi desenvolvida por empresas do norte. Desse modo, desconsideram-se, em muitos casos, aspectos específicos das culturas do sul, até porque os conjuntos de dados mais populares são centrados nos Estados Unidos e na Europa ocidental. Nesse sentido, é comum que um sistema de reconhecimento de imagem classifique uma fotografia de uma mulher em um vestido branco como uma noiva, mas não o faça com a imagem de uma mulher trajando um sári na celebração de seu casamento (Cortiz, 2020).

Tudo isso aumenta o risco de as crianças serem identificadas ou até mesmo acusadas erroneamente, sendo que as consequências não desejadas têm o potencial de gerar repercussões duradouras por toda a vida delas. Tais consequências podem se dar em diversas esferas, como na sua educação e em oportunidades de emprego quando adultas, além de poder causar um impacto relevante no seu comportamento e desenvolvimento, em especial, para as crianças integrantes de grupos mais vulneráveis.

A American Civil Liberties Union (ACLU), nesse sentido, suspeita que os sistemas de reconhecimento facial, nas escolas dos Estados Unidos, tenham como alvo estudantes negros por mau comportamento, reforçando, assim, a criminalização deste grupo identitário, historicamente, discriminado com base na sua raça (UC Berkeley Human Rights Center Research Team, 2019).

Outra fonte de discriminação levantada por conjuntos de dados pobres é a discriminação baseada no gênero porque o reconhecimento facial, em geral, é menos preciso para mulheres e meninas, especialmente as ne-

20 Disponível em: https://www.technologyreview.com/2020/10/09/1009992/live-facial-recognition-is-tracking-kids-suspected-of-crime/. Acesso em: 17 set. 2021.

gras, na medida em que algoritmos têm desempenho pior para rostos de mulheres do que para rostos masculinos (Grother; Ngan; Hanaoka, 2019).

A respeito de discriminação algorítmica contra meninas, há, também, o notório caso envolvendo a rede social Instagram, que foi acusada de impulsionar conteúdos de padrão corporal a ensejar danos psicológicos e de saúde mental em garotas adolescentes[21]. Cumpre ressaltar, a propósito, que é durante a fase da inicial da adolescência que se desenvolvem capacidades cerebrais fundamentais para o ser humano conseguir resistir a impulsos que lhe são estimulados por fatores externos, o que potencializa a vulnerabilidade das pessoas nessa fase de vida (Unicef, 2021).

Vale ainda mencionar o recente exemplo de discriminação algorítmica, igualmente, com potencial de danos para toda a vida adulta, que se deu com o sistema de IA usado pelo Reino Unido, durante a pandemia, para avaliar e classificar jovens estudantes ao ingresso nas universidades daquele país. O sistema ignorou talentos individuais e considerou o coletivo das escolas, rebaixando as notas de estudantes excelentes de escolas de baixo desempenho. Com isso, estudantes de escolas privadas acabaram sendo beneficiados e estudantes mais pobres e negros, prejudicados. A utilização desse algoritmo motivou diversas manifestações contra o algoritmo usado – talvez as primeiras manifestações públicas contra um algoritmo![22]

São, com efeito, inúmeras as circunstâncias que geram discriminação no ambiente *online*, podendo-se dizer que, dentre as mais recorrentes, estão as discriminações raciais, de gênero e relacionadas às desigualdades socioeconômicas (Eubanks, 2019), que exacerbam as desigualdades já presentes em vários contextos sociais, além de colocar em risco a própria democracia (O'Neil, 2016).

No que diz respeito às crianças, imperioso notar que elas sofrem discriminação e inequidade racial, de gênero ou por condição econômica, de forma interseccional nesses diferentes grupos sociais. Além disso, vivenciam, simultaneamente, uma das dinâmicas de poder sociais mais desiguais e mesmo violentas que ainda persistem nas sociedades contemporâneas: as relações adultocêntricas. Por meio da naturalização dessa relação hierar-

21 Disponível em: https://www.wsj.com/articles/facebook-knows-instagram-is-toxic-for-teen-girls-company-documents-show-11631620739. Acesso em: 26 nov. 2021.

22 Disponível em: https://g1.globo.com/mundo/noticia/2020/08/20/algoritmo-roubou-meu-futuro-solucao-para-enem-britanico-na-pandemia-provoca-escandalo.ghtml. Acesso em: 17 set. 2021.

quizada, crianças tornaram-se objeto de exploração e abuso por diversas instituições e pessoas, inclusive familiares, enfrentando uma opressão social única (Bustelo, 2007).

É, pois, imprescindível que sistemas de IA sejam cuidados, também sob uma perspectiva inclusiva, para que as pessoas e, notadamente, as crianças possam deles usufruir adequada e sadiamente, na sua maior potência, estando protegidas *no* ambiente digital e não *do* ambiente digital (Denham, 2019).

2 DIREITOS FUNDAMENTAIS DE CRIANÇAS E A RELAÇÃO ENTRE PRINCÍPIOS ÉTICOS ASSOCIADOS À IA E PRINCÍPIOS E CONCEITOS DA PROTEÇÃO DE DADOS PESSOAIS

O direito à igualdade e à não discriminação está previsto, internacionalmente, no sistema global de proteção aos direitos humanos. Está presente no art. 7º da Declaração Universal dos Direitos Humanos, de 1948[23], e no art. 26 do Pacto Internacional sobre Direitos Civis e Políticos, de 1966[24], que, promulgado pelo Decreto nº 592/1992[25], traz, ainda, uma especial atenção ao direito da criança à não discriminação, no seu art. 24. Também na Convenção sobre a Eliminação de Todas as Formas de Discriminação Racial, promulgada no Brasil pelo Decreto nº 65.810/1969[26], e na Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher, de 1979, promulgada pelo Decreto nº 4.377/2002[27].

No Brasil, o direito à não discriminação é garantido pela Constituição Federal, a qual, no *caput* do art. 5º, prevê que todas as pessoas são iguais perante a lei, sem qualquer distinção. Especificamente sobre as crianças, o País promulgou, pelo Decreto nº 99.710/1990[28], a Convenção sobre os Direitos da Criança, que, logo no seu art. 2º, menciona que os países respeitarão os direitos da criança sem qualquer discriminação, independentemente "de raça, cor, sexo, idioma, crença, opinião política ou de outra

23 Disponível em: https://www.un.org/en/about-us/universal-declaration-of-human-rights. Acesso em: 18 set. 2021.

24 Disponível em https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx. Acesso em: 18 set. 2021.

25 Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/1990-1994/d0592.htm. Acesso em 18 set. 2021.

26 Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/1950-1969/D65810.html. Acesso em: 18 set. 2021.

27 Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/2002/d4377.htm. Acesso em: 18 set. 2021.

28 Disponível em: http://www.planalto.gov.br/ccivil_03/decreto/1990-1994/d99710.htm. Acesso em: 18 set. 2021.

índole, origem nacional, étnica ou social, posição econômica, deficiências físicas, nascimento ou qualquer outra condição da criança, de seus pais ou de seus representantes legais". Também o Estatuto da Criança e do Adolescente prevê, no art. 3º, que os direitos das crianças aplicam-se a todas elas, indistintamente, sem quaisquer discriminação, bem como, diz no art. 5º, que nenhuma criança será objeto de qualquer forma de discriminação.

A criança não é um miniadulto; é sujeito de direitos que vivencia um estágio peculiar de desenvolvimento e, como tal, deve ter seus direitos humanos e fundamentais garantidos em todas as esferas da sua vida. Nesse sentido, o art. 227 da Constituição Federal não deixa dúvidas sobre seus direitos fundamentais, como o direito a uma vida com dignidade, bem como ao respeito, ao lazer, à convivência familiar e comunitária, entre outros. E mais: determina que o Estado, a sociedade e as famílias têm o dever de garantir tais direitos com prioridade absoluta, de forma que a criança seja cuidada com primazia, além de colocá-las "a salvo de toda forma de negligência, discriminação, exploração, violência, crueldade e opressão".

Em relação aos direitos da criança no ambiente digital, o recente Comentário Geral nº 25 (ONU, 2021) é contundente ao apresentar o princípio da não discriminação como um dos quatro que o orientam – ao lado do melhor interesse; do direito à vida, à sobrevivência e ao desenvolvimento; e ao respeito pela opinião da criança –, sendo parte essencial à implementação dos direitos humanos das crianças nesse ambiente, inclusive em relação à IA.

Com relação à IA, cumpre dizer que o Brasil possui uma estratégia nacional a esse respeito, que foi instituída pela Portaria nº 4.617/2021[29-30]. Nessa estratégia, entre outros tópicos, é mencionada a recomendação da OECD (OECD, 2019) sobre IA, à qual o Brasil aderiu e que apresenta alguns elementos, como: a importância de que a IA esteja a serviço do ser humano, beneficiando as pessoas e o planeta, bem como impulsionando o crescimento inclusivo, o desenvolvimento sustentável e o bem-estar; de forma que seus respectivos sistemas sejam projetados com respeito ao Estado de Direito, aos direitos humanos, aos valores democráticos e à diversidade.

29 Disponível em: https://www.in.gov.br/en/web/dou/-/portaria-gm-n-4.617-de-6-de-abril-de-2021-*-313212172. Acesso em: 19 set. 2021.

30 Vale noticiar, ainda, o PL 21/2020, que intenta regulamentar a IA no país e, sem as necessárias discussões, foi aprovado pela Câmara dos Deputados, mas segue pendente de análise pelo Senado.

A Estratégia Brasileira para a IA, disposta na citada Portaria, está fundada em cinco princípios, dentre os quais se encontram os valores centrados no ser humano e na equidade; a transparência e explicabilidade; a robustez, segurança e proteção e a responsabilização ou a prestação de contas (*accountability*).

Esses princípios são de suma importância para toda a discussão acerca da discriminação algorítmica, porquanto os sistemas de IA têm sido utilizados, com a tomada de decisões automatizadas, sem que se saiba se tais decisões são mesmo apropriadas (Frazão, 2021). Isso porque, como já assinalado neste artigo, sistemas de IA operam com base no reconhecimento de padrões, fazendo correlações e realizando inferências. Em decorrência da sua capacidade de aprendizado desenvolvido na relação com a base de dados que alimentam seus algoritmos, aprendem a fazer generalizações, predições e categorizações, "solucionando problemas de maneira dinâmica, ainda que não tenham sido programados para tanto" (Wimmer, 2019, p. 383).

Essas características dos sistemas de IA suscitam inúmeras questões, sendo que, do ponto de vista da responsabilidade jurídico-legal, é justamente a opacidade dos processos decisórios um elemento central de discussão (Wimmer, 2019). Mesmo porque o avanço tecnológico no desenvolvimento e na implementação de sistemas de IA não pode dispensar o fator humano e a necessidade de os sistemas estarem a serviço do ser humano, atuando eticamente, inclusive no que diz respeito à tomada de decisão automatizada (Santaella, 2021), em especial quando se trata de crianças, que são, sabidamente, mais vulneráveis.

Daí a necessidade de que, em consonância com os princípios comuns da equidade, confiabilidade, segurança e responsabilidade, bem como da necessidade de que os sistemas de IA sejam centrados no humano, conforme estabelecido em inúmeros documentos internacionais sobre o tema (Burle; Cortiz, 2019), haja uma calibragem entre a auditabilidade, a transparência e a explicabilidade dos sistemas de IA. Esses princípios também são fundamentais para uma IA que garanta os direitos de crianças e, com isso, favoreça a sua efetiva participação e compreensão do funcionamento dos sistemas (Unicef, 2020).

Do mesmo modo, é fundamental estabelecer parâmetros que possam ser considerados para uma adequada apuração e definição de responsabilidades no caso de infrações cometidas com base nestes sistemas (Wimmer,

2019), e também no que diz respeito às resultantes discriminatórias e preconceituosas das decisões algorítmicas.

Como bem assevera a Relatora Especial das Nações Unidas sobre racismo, discriminação racial, xenofobia e intolerância relacionada, Tendayi Achiume, ao citar a Convenção sobre a Eliminação de Todas as Formas de Discriminação Racial, é imperioso que sejam asseguradas a reparação e a compensação por danos sofridos como resultado de discriminação racial no ambiente digital, assim como a prevenção e a mitigação de discriminação em sistemas de IA, inclusive por *due diligence* em direitos humanos por parte das empresas envolvidas (ONU, 2020).

Em razão de os sistemas de IA serem desenvolvidos a partir da capacidade de tratamento de bases de dados em velocidade, volume e variedade sem precedentes, é certo que, para fins da garantia dos direitos de todas as pessoas, também de crianças e adolescentes, a LGPD é de suma importância. Dados pessoais são extensões da personalidade e da própria pessoa (Doneda, 2020), sendo a sua proteção direito fundamental do indivíduo (Sarlet, 2021).

Por isso, os princípios basilares da LGPD para o tratamento de dados pessoais – finalidade, adequação, necessidade, livre acesso, qualidade dos dados, transparência, segurança, prevenção, não discriminação, responsabilização e prestação de contas – são relevantes para a discussão acerca da discriminação resultante de decisões das máquinas em sistemas de IA, em especial quando se trata de crianças. O uso indiscriminado de dados pessoais é capaz de objetificar as pessoas, promover manipulações, afetar o livre desenvolvimento da personalidade e gerar discriminações. Da mesma forma, prática corrente em sistemas de IA, a criação de perfis, por meio de perfilamento das pessoas e criação de parâmetros de avaliação sobre aspectos da personalidade dos indivíduos, tem grande potencial de criar discriminações, as quais, no caso de crianças, possuem alto risco de acarretar impactos prejudiciais com reflexos por toda a sua vida.

Ao promover discriminação em face do titular dos dados pessoais ou em relação ao grupo social que representa, sob o aspecto racial, de gênero, etário, relacionado a ter ou não deficiências, o resultado da aplicação do algoritmo, além de violar princípios éticos da IA e o direito constitucional à igualdade, deixa, também, de promover, adequadamente, o direito fundamental à proteção de dados. No caso da proteção de dados, é importante lembrar aqui os usos de IA em brinquedos, como é o caso da boneca Cayla,

proibida na Alemanha, ou da *smart* tv, que recolhem dados em ambientes domésticos, ou ainda o processo de recolha de dados feito pelas plataformas com fins de impulsionamento de conteúdos, sem que haja transparência quanto ao seu uso.

No Comentário Geral nº 25 (ONU, 2021), que, no seu processo de elaboração, ouviu crianças de 28 países, elas evidenciaram em seus depoimentos o desejo de saber mais sobre o destino dos seus dados. Trata-se de uma questão crucial, como reconhece o documento, associada ao direito da criança de buscar, receber e difundir informação em um ambiente seguro.

Se for considerado o outro lado da coleta de dados, ou seja, o impulsionamento por meio do uso de IA, resta evidente que se trata do próprio acesso à cultura, em sua universalidade e diversidade, que fica comprometido quando, no lazer infantil, a criança recebe conteúdos que foram impulsionados por uma lógica mercadológica afeita a outros interesses que não o seu melhor interesse. Qualquer criança que acesse as redes sociais no país não conseguirá fazê-lo de forma absolutamente livre do assédio comercial. É estimulada a permanecer conectada, curtir e compartilhar conteúdos, sendo monitorada em suas pegadas digitais, passando a receber recomendações que, muitas vezes, contribuem – de modo intencional ou não – para polarizações, discriminações e exclusão.

Em *playlists* como criança rica *v.* criança pobre; meninos *v.* meninas, entre outros, sob o argumento do que se trata de uma brincadeira, massifica-se a discriminação (Sampaio; Pereira; Cavalcante, 2021). Nas imagens da abundância para se ver e desejar, a casa vira uma mansão, o cenário da ficção se transforma num *shopping*, a popularidade em alta vale mais que a amizade. O impulsionamento altera, portanto, profundamente o acesso ao repertório cultural das crianças e o horizonte de sua produção. É crucial que essa lógica seja revista, a começar pelo atendimento do princípio da transparência.

No caso de crianças, em razão do alto risco envolvido quanto às consequências potencialmente danosas, posto que para toda a vida, advindas de um abusivo ou inadequado tratamento de dados pessoais, é imprescindível seja considerada a proteção jurídica especial que as assiste (Henriques; Pita; Hartung, 2021).

Nesse sentido, a determinação de que o tratamento de dados pessoais de crianças seja feito, exclusivamente, se for em seu "melhor interesse", nos termos do *caput* do art. 14, coaduna-se, por completo, com o princí-

pio essencial de que sistemas de IA que as afetem estejam nelas centrados (Unicef, 2020). Não é possível, por exemplo, que o interesse do controlador ou de terceiros – ainda que legítimo – seja utilizado como base legal para o tratamento de dados de crianças, porquanto é o melhor interesse delas que, sempre, deverá prevalecer (Henriques, 2021).

A aplicação do Direito e da LGPD é, pois, fundamental para que realidades estruturadas com base em sistemas de dominação (Moreira, 2020) sejam coibidas de manter ou exacerbar as disparidades entre grupos sociais também no âmbito do uso de algoritmos de tomada de decisão. Trata-se de um movimento que deve seguir, paralelamente, ao esforço de aplicação das normas de direitos humanos e fundamentais, para além do ambiente digital, com vistas a que as próprias sociedades sejam mais inclusivas e antidiscriminatórias.

De qualquer forma, é certo que só o Direito não dará conta desse imenso desafio. Os caminhos para que sistemas de IA sejam inclusivos e não discriminatórios passam, também e concomitantemente, pelo próprio desenvolvimento ético destes sistemas, pela atuação responsável das empresas e profissionais envolvidos e pelas normas sociais.

3 CAMINHOS PARA A SOLUÇÃO

A inclusão social é o oposto da exclusão social e da própria discriminação. Pode ser tratada como um conceito relacional, por meio do qual indivíduos ou grupos são incluídos com outros indivíduos, em outros grupos e na sociedade de maneira geral. Pode ser vista do prisma contextual, ligada a questões estruturais e afetada por dimensões não só locais, mas regionais, nacionais e mesmo globais. Pode, ainda, estar relacionada a instituições que criam estruturas diversas e podem, ou não, reproduzir e reforçar desigualdades históricas. Um mesmo indivíduo pode pertencer a várias identidades e, com elas, ser incluído ou estar sujeito à exclusão, podendo enfrentar uma opressão social única pelo acúmulo destas identidades e pelo fato de a exclusão social poder ser um conjunto de processos simultâneos e dinâmicos, ao invés de uma condição fixa.

Não existem fórmulas prontas para tornar éticos e inclusivos os sistemas de IA. Contudo, os caminhos para a solução do problema da discriminação algorítmica, indubitavelmente, passam pela necessidade de colaboração proativa entre cientistas de dados, sociedade civil, formuladores de políticas, governos, setor privado, investidores e especialistas, em uma

abordagem multissetorial, com uma maior participação, quiçá igualitária, de mulheres e pessoas negras em todas essas áreas. Mas nada disso será suficiente, caso não seja assegurado um espaço de escuta e participação das crianças que considere, efetivamente, suas peculiaridades de pessoas em desenvolvimento, o que implica o uso de linguagem, dinâmica e tempo ajustados às suas necessidades.

Da mesma maneira, é fundamental contar com uma equipe heterogênea – com diferentes etnias, origens, religiões, gêneros e raças – e interdisciplinar tanto para a criação e o desenvolvimento de sistemas de IA, como para se discutir temas relacionados à ética, equidade e justiça junto aos desenvolvedores. Nesse sentido, é essencial que haja um amplo esforço para uma profícua educação sobre valores humanos para a IA.

Os princípios éticos da IA, centrada no ser humano, também devem pautar a conduta de todos aqueles que fazem parte da cadeia de desenvolvimento dos sistemas de IA: auditabilidade, prestação de contas, explicabilidade, justiça e transparência. De maneira a atender aos direitos humanos, tais princípios devem orientar as empresas e suas práticas de autorregulação e *compliance*; as leis e órgãos públicos fiscalizadores, o *design* de sistemas de IA e a própria sociedade. E, no caso de crianças, tais princípios orientadores devem, ainda, estar conectados com o dever de garantia do melhor interesse desse grupo de pessoas vulnerável.

Por fim, é também fundamental garantir que indivíduos e comunidades diretamente impactados por uma tecnologia específica de IA possam influenciar o seu desenvolvimento – participando do desenho, do teste e da auditoria –, não sendo, meramente, relegados à condição de participantes e usuários passivos de novos sistemas de IA (Rendtorff, 2018).

Essa participação deve ser assegurada, com especial atenção, a residentes em países do sul global e em comunidades mais vulneráveis, a fim de que sejam desenvolvidas soluções também para mercados nos quais as pessoas não são consideradas "alfabetizadas em dados" e "digitalmente capazes". Isso na comparação com aquelas que participam dos "mercados disponíveis" em comunidades mais abastadas e com acesso à educação formal, de maneira que o hiato entre tais mercados não venha a ser tão grande a, futuramente, inviabilizar investimentos no espaço da tecnologia e inovação, exacerbando, sobremaneira, as desigualdades e fomentando discriminações preconceituosas (Nwakanma, 2020).

No que diz respeito às crianças, a participação delas é igualmente desejável, notadamente em relação ao desenvolvimento de sistemas de IA que possam impactá-las diretamente, em especial nas realidades mais vulneráveis. O desafio aqui é criar mecanismos para assegurar esse processo de escuta desde o desenho dos sistemas de IA ao seu processo de implementação e avaliação. Trata-se de conceber, em termos éticos, os parâmetros desse processo de participação, inclusive com a definição de protocolos de acesso público.

Vale ressaltar que, hoje, tem-se conhecimento que empresas de *ad-tech* coletam 72 milhões de pontos de dados sobre uma criança até ela chegar aos 13 anos de idade (Global Action Plan, 2020). Trata-se de dados que alimentarão sistemas de IA diversos, sem que, muitas vezes, crianças e até mesmo adultos tenham dele e de suas implicações qualquer conhecimento.

Daí por que a IA que impacta as múltiplas infâncias deve estar centrada na criança, bem como a governança de dados pessoais de crianças também deve, igualmente, estar nelas centrada (Unicef, 2021), de forma a ser garantido o melhor interesse e os direitos humanos de todas as crianças, sem qualquer discriminação étnico-racial, socioeconômica, de gênero, por condição de deficiência ou outra qualquer.

Exemplos positivos do uso de sistemas de IA nessa direção atestam a possibilidade efetiva de que estejam a serviço dos direitos humanos das crianças, auxiliando-as a desenvolverem todo o seu potencial. Os próprios algoritmos, aliás, podem ser usados para detectar e combater a discriminação – como no caso do *chatbot* da Unicef criado para enfrentar discriminação contra crianças venezuelanas no Brasil[31]. Da mesma forma, o caso da criança autista que desenvolveu habilidades de linguagem conversando com a Siri[32] (Newman, 2016) mostra que as potencialidades são enormes e que há um caminho possível pela frente. Por outro lado, é também crucial reconhecer que muito mais pode e deve ser feito.

31 Disponível em: https://www.unicef.org/brazil/comunicados-de-imprensa/unicef-lanca-chatbot-para-enfrentar-discriminacao-contra-criancas-e-adolescentes-venezuelanos-no-brasil. Acesso em: 20 set. 2021.

32 Sem desconsiderar as críticas aos assistentes pessoais, inclusive, no que diz respeito à discriminação de gênero, por serem, na sua maioria, vozes femininas ambientadas em uma posição servil.

CONSIDERAÇÕES FINAIS

A IA, pautada no ser humano e, por conseguinte, nos valores humanos (LI, 2018), pode contribuir, positivamente, em muitos aspectos da vida de todas as pessoas, inclusive de crianças nas múltiplas infâncias existentes ao redor do mundo. Para isso, é fundamental que esteja ancorada na garantia dos direitos humanos e na proteção, provisão e participação de todas as crianças, sem discriminações preconceituosas, em um mundo digital que seja projetado com as crianças, para que possam acessá-lo de forma criativa, com conhecimento e sem medo.

REFERÊNCIAS

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine Bias: there's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*. 2016. Disponível em: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Acesso em: 17 set. 2021.

ANISTIA INTERNACIONAL; ACCESS NOW. *Declaração de Toronto*. Toronto: Anistia Internacional e Access Now, 2018. Disponível em: https://www.torontodeclaration.org/declaration-text/english/. Acesso em: 8 set. 2021.

BAUMAN, Zygmunt; LYON, David. *Vigilância líquida*. Rio de Janeiro: Zahar, 2013.

BIGONHA, Carolina. #TechforGood. In: Comitê Gestor da Internet no Brasil (CGI). Inteligência Artificial em perspectiva. *Panorama Setorial da Internet*, n. 2, a. 10, 2018.

BUSTELO, Eduardo. *El recreo de la infância*: argumentos para outros comienzos. Buenos Aires: Siglo Veintiuno Editores, 2007.

BURLE, Caroline; CORTIZ, Diogo. *Mapeamento de princípios de inteligência artificial*. São Paulo: CEWEB.BR, 2020 [livro eletrônico]. Disponível em: https://ceweb.br/media/docs/publicacoes/17/20200721143359/digital_mapeamento_principios_IA_portugues.pdf. Acesso em: 27 ago. 2021.

CENTRO REGIONAL DE ESTUDOS PARA O DESENVOLVIMENTO DA SOCIEDADE DA INFORMAÇÃO (CETIC.BR). *Pesquisa Tic Kids Online Brasil 2019*. São Paulo: CETIC.BR, 2019.

CORMEN, Thomas H. *Algorithms unlocked*. Cambridge: MIT Press, 2013.

CORTIZ, Diogo. Inteligência artificial: equidade, justiça e consequências. In: Comitê Gestor da Internet no Brasil (CGI). *Panorama Setorial da Internet*, n. 1, a. 12, 2020.

DENHAM, Elizabeth. *Protecting children online*: update on progress of ICO code. A blog by Elizabeth Denham, Information Commissioner, 2019. Disponível em: https://ico.org.uk/about-the-ico/news-and-events/blog-protecting-children-online-update-on-progress-of-ico-code/. Acesso em: 10 set. 2021.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. São Paulo: Thomson Reuters Revista dos Tribunais, 2020.

_____; ALMEIDA, Virgilio. *O que é a governança de algoritmos?* Disponível em: https://politics.org.br/edicoes/o-que-%C3%A9-governan%C3%A7a-de-algoritmos. Acesso em: 5 set. 2021.

_____; MENDES, Laura Schertel; SOUZA, Carlos Affonso Pereira de; ANDRADE, Norberto Nuno Gomes. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. UNIFOR – Universidade de Fortaleza. *Revista de Ciências Jurídicas Pensar*, v. 23, n. 4, 2018. Disponível em: https://periodicos.unifor.br/rpen/article/view/8257/pdf#. Acesso em: 3 set. 2021.

ECPAT INTERNATIONAL. *Regional overview*: sexual exploitation of children in the middle east and north Africa. Bangkok: ECPAT International, 2020.

EUBANKS, Virginia. *Automating inequality*: how high-tech tools profile, police, and punish the poor. Nova Iorque: Picador, 2019.

FERRARI, Isabela. Entrevista. In: Comitê Gestor da Internet no Brasil (CGI). *Panorama Setorial da Internet*, n. 1, a. 12, 2020.

FUNDAÇÃO ABRINQ. *Cenário da infância e adolescência no Brasil*. São Paulo: Fundação Abrinq, 2021.

FRAZÃO, Ana. Discriminação algorítmica: o hiato entre quem programa e quem usa – A terceirização de processos decisórios por agentes públicos e privados. Parte IV. *Jota*, 2021. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-o-hiato-entre-quem-programa-e-quem-usa-07072021. Acesso em: 20 set. 2021.

GARDNER, Howard. *O verdadeiro, o belo e o bom redefinidos*: novas diretrizes para a educação no século XXI. Rio de Janeiro: Rocco, 2012.

GETSCHKO, Demi. *Inteligência artificial e IoT*. Fórum Brasileiro de Internet das Coisas. 2021. Disponível em: https://www.youtube.com/watch?v=_17QAD7ujH4&t=3667s. Acesso em: 20 set. 2021.

GLOBAL ACTION PLAN. *Global Action Plan joins forces with campaigners to demand Google and major tech firms end targeted ads to children*, 2020. Disponível em: https://www.globalactionplan.org.uk/news/global-action-plan-joins-forces-with-campaigners-to-demand-google-and-major-tech-firms-end-targeted-ads-to-children. Acesso em: 20 set. 2021.

GROTHER, Patrick; NGAN, Mei; HANAOKA, Kayee. *Face recognition vendor test (FRVT) Part 3*: demographic effecs. National Institute of Standards and Technology.

U. S. Department of Commerce. 2019. Disponível em: https://doi.org/10.6028/NIST.IR.8280. Acesso em: 17 set. 2021.

HARARI, Yuval Noah. *Homo Deus*: uma breve história do amanhã. São Paulo: Companhia das Letras, 2016.

HENRIQUES, Isabella. Inteligência artificial e a nova economia de dados: reflexões na perspectiva da infância brasileira. In: CANTARINI, Paola; GUERRA FILHO, Willis Santiago; KNOERR, Viviane Coêlho de Séllos. *Direito e inteligência artificial*: Fundamentos. Volume 2: Inteligência artificial e tutela de direitos. Rio de Janeiro: Lumen Juris, p. 111-144, 2021.

HENRIQUES, Isabella; PITA, Marina; HARTUNG, Pedro. A proteção de dados pessoais de crianças e adolescentes. In: MENDES, Laura Schertel; DONEDA, Danilo; SARLET, Ingo Wolfgang; RODRIGUES JR., Otavio Luiz. *Tratado de proteção de dados pessoais*. Rio de Janeiro: Forense, p. 199-225, 2021.

INTERNATIONAL TELECOMMUNICATION UNION (ITU). *Measuring digital development*: facts and figures 2020. Genebra: ITU Publications. 2020. Disponível em: https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf. Acesso em: 10 set 2021.

ITS; BERKMAN KLEIN CENTER; GLOBAL NETWORK OF INTERNET AND SOCIETY RESEARCH CENTERS. *Artificial intelligence & inclusion*. 2017. Disponível em: https://aiandinclusion.org/#home Acesso em: 20 set. 2021.

LESSIG, Lawrence. The law of the horse: what cyberlaw might teach. *Harvard Law Review*, v. 113, n. 2, 1999, p. 501-546. Disponível em: https://cyber.harvard.edu/works/lessig/finalhls.pdf. Acesso em: 7 set. 2021.

LI, Fei-Fei. *Machine values are human values*. New Work Summit by The New York Times. 2018. Disponível em: https://www.nytimes.com/video/admin/100000005753299/li-machine-values-are-human-values.html. Acesso em: 21 set. 2021.

LIPOVETSKY, Gilles. *A era do vazio*: ensaios sobre o individualismo contemporâneo. Barueri: Manole, 2005.

MARINO, Eduardo; CHICARO, Marina Fragata. FMCSV, TJSP e Alana: uma parceria promotora do desenvolvimento da primeira infância. In: HENRIQUES, Isabella (Org.). *Primeira infância no sistema de garantia de direitos de crianças e adolescentes* – Uma experiência a ser replicada. São Paulo: Instituto Alana, 2019.

MENDES, Laura; MATTIUZZO, Marcela. Discriminação algorítmica: conceito, fundamento legal e tipologia. Brasília: *Revista Direito Público*, v. 16, n. 90, 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766. Acesso em: 8 set. 2021.

MOREIRA, Adilson José. *Tratado de direito antidiscriminatório*. São Paulo: Contracorrente, 2020.

MULHOLLAND, Caitlin; FRAJHOF, Isabella. Inteligência artificial e a Lei Geral de Proteção de Dados Pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de *machine learning*. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. *Inteligência artificial e Direito* – Ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, p. 265-290, 2019.

NEWMAN, Judith. *Siri with love*: a mother, her autistic son, and the kindness of machines. Londres: Quercus, 2016.

NWAKANMA, Nnenna. Entrevista. In: Comitê Gestor da Internet no Brasil (CGI). *Panorama Setorial da Internet*, n. 1, a. 12, 2020.

OECD. Council Recommendation on Artificial Intelligence. 2019. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 20 set. 2021.

O'NEIL, Cathy. *Weapons of math destruction*: how big data increases inequality and threatens democracy. Nova Iorque: Crown, 2016.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). *Racial discrimination and emerging digital technologies*: a human rights analysis. 2020. Disponível em: https://undocs.org/en/A/HRC/44/57. Acesso em: 7 set. 2021.

_____. *Comentário Geral nº 25 sobre os direitos da criança em relação ao ambiente digital*. Comitê dos Direitos da Criança da ONU, 2021. Disponível em: https://criancaeconsumo.org.br/biblioteca/comentario-geral-n-25/. Acesso em: 20 set. 2021.

ORGANIZATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). *O elevador social está quebrado?* Como promover a mobilidade social. 2018. Disponível em: https://www.oecd.org/brazil/social-mobililty-2018-BRA-PT.pdf. Acesso em: 17 set. 2021.

PIAGET, Jean; INHELDER, Barbel. *A psicologia da criança*. Trad. Octavio Mendes Cajado. 11. ed. Rio de Janeiro: Difel, 2021.

RENDTORFF, Sara. Entrevista. In: Comitê Gestor da Internet no Brasil (CGI). *Inteligência artificial em perspectiva*.

SAMPAIO, Inês Vitorino; PEREIRA, Georgia C.; CAVALCANTE, Andrea P. P. Crianças youtubers e o exercício do direito à comunicação. *Revista Cedes*, Campinas, v. 41, n. 113, p. 14-22, jan./abr. 2021. Disponível em: https://www.scielo.br/j/ccedes/a/3sMFJ336TSHB4fzg3XNyFJr/?format=pdf&lang=pt. Acesso em: 29 nov. 2021.

SANTAELLA, Lucia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis Santiago; SANTAELLA, Lucia; KAUFMAN, Dora; CANTARINI, Paola. *Direito e inteligência artificial*: fundamentos. Volume 1: Inteligência artificial, ética e Direito. Rio de Janeiro: Lumen Juris, p. 109-136, 2021.

SARLET, Ingo Wolfgang. Fundamentos constitucionais: o direito fundamental à proteção de dados. In: MENDES, Laura Schertel; DONEDA, Danilo; SARLET, Ingo Wolfgang; RODRIGUES JR., Otavio Luiz. *Tratado de proteção de dados pessoais*. Rio de Janeiro: Forense, p. 21-59, 2021.

SAS INSTITUTE. *Artificial Intelligence* – What it is and why it matters. Disponível em: https://www.sas.com/en_us/insights/analytics/what-is-artificial-intelligence.html. Acesso em: 5 set. 2021.

SILVA, Tarcízio. *Linha do tempo do racismo algorítmico*. Blog do Tarcízio Silva. 2021. Disponível em: http://https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo. Acesso em: 4 set. 2021.

SOUZA, Carlos Affonso Pereira de; OLIVEIRA, Jordan Vinícius de. Sobre os ombros de robôs? A inteligência artificial entre fascínios e desilusões. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. *Inteligência artificial e Direito* – Ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, p. 65-81, 2019.

THE ALAN TURING INSTITUTE. *Frequently asked questions*. Disponível em: https://www.turing.ac.uk/about-us/frequently-asked-questions. Acesso em: 5 set. 2021.

TURNER, Jacob. *Robot rules*: regulating artificial intelligence. Nova Iorque: Palgrave Macmillan, 2019.

UC BERKELEY HUMAN RIGHTS CENTER RESEARCH TEAM. *Memorandum on Artificial Intelligence and Child Rights*. 2019. Disponível em: https://www.unicef.org/innovation/reports/memoAIchildrights. Acesso em: 17 set. 2021.

UNITED NATIONS CHILDREN'S FUND (UNICEF). *The adolescent brain*: a second window of opportunity. 2017. Disponível em: https://www.unicef-irc.org/publications/933-the-adolescent-brain-a-second-window-of-opportunity-a-compendium.html. Acesso em: 26 nov. 2021.

_____. *The Case for Better Governance of Children's Data*: a Manifesto. 2021. Disponível em: https://www.unicef.org/globalinsight/reports/better-governance-childrens-data-manifesto. Acesso em: 20 set. 2021.

_____. *Policy guidance on AI for children*. 2020. Disponível em: https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf. Acesso em: 5 set. 2021.

_____. *Children in a digital world* – The state of the world's children. 2017. Disponível em: https://www.unicef.org/media/48601/file. Acesso em: 3 set. 2021.

WEST, Sarah Myers; WHITTAKER, Meredith; CRAWFORD, Kate. *Discriminating systems*: gender, race and power in AI. Nova Iorque: New York University, AI Now Institute, 2019. Disponível em: https://ainowinstitute.org/discriminatingsystems.pdf. Acesso em: 8 set. 2021.

WIMMER, Miriam. Responsabilidade de agentes empresariais por ilícitos administrativos praticados por sistemas de inteligência artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin. *Inteligência artificial e Direito* – Ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, p. 373-395, 2019.

ZUBOFF, Shoshana. *The age of surveillance capitalism*: the fight for a human future at the new frontier of power. Nova Iorque: Public Affairs, 2020.

Sobre as autoras:

Isabella Vieira Machado Henriques | *E-mail:* isahenriques@hotmail.com

Doutoranda em Direito das Relações Sociais – Direitos Difusos e Coletivos – pela Pontifícia Universidade Católica de São Paulo (PUC/SP). Mestra em Direito pela PUC/SP. Advogada. Diretora Executiva do Instituto Alana. Presidente da Comissão de Defesa dos Direitos da Criança e do Adolescente da OAB/SP. Conselheira do Conselho Consultivo da Ouvidoria da Defensoria Pública do Estado de São Paulo. Conselheira e Cofundadora do Advocacy Hub. Autora de livros e artigos sobre temas relacionados a direitos fundamentais e crianças. Pesquisadora sobre temas de proteção de dados pessoais e direitos das crianças.

Inês Vitorino Sampaio | *E-mail:* inesvict@gmail.com

Doutora em Ciências Sociais pela Universidade Estadual de Campinas (Unicamp) e Mestre em Sociologia pela Universidade Federal do Ceará (UFC). Docente do Programa de Pós-Graduação em Comunicação da UFC e Vice-Coordenadora do Laboratório de Pesquisa da Relação Infância, Juventude e Mídia (LabGRIM). Tem como principais interesses de pesquisa a relação de crianças e adolescentes com a comunicação sob a ótica dos direitos. Autora do livro *Televisão, publicidade e infância* (2004), entre outras publicações.

Data de submissão: 30 de setembro de 2021.
Data de aceite: 10 de janeiro de 2022.

A Inteligência Artificial no Contexto Atual: uma Análise à Luz das Neurociências Voltada para uma Proposta de Emolduramento Ético e Jurídico

Artificial Intelligence in the Current Context: an Analysis in the Light of the Neurosciences Aimed at a Proposal for an Ethical and Legal Framework

GABRIELLE BEZERRA SALES SARLET[1]

Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS).

RESUMO: Examina-se o uso irreflexivo/abusivo de tecnologias disruptivas e pervasivas como a inteligência artificial para, partindo da atual contribuição das neurociências, especialmente no que toca à função da emocionalidade no processo decisório, apontar o novo sentido e o alcance das vulnerabilidades na sociedade informacional e, nesse sentido, esboçar algumas pautas para a sua regulação, para além da proteção de dados pessoais, e, portanto, uma conformação com a proteção multinível da pessoa humana dentro e fora do ecossistema virtual/digital. O método científico envolve uma abordagem hipotético-dedutiva, com pesquisa explicativa, exploratória e procedimento bibliográfico.

PALAVRAS-CHAVE: Inteligência artificial; neurociências; direitos humanos e fundamentais; sociedade informacional; proteção de dados pessoais.

ABSTRACT: The irreflexive/abusive use of disruptive and pervasive technologies, such as artificial intelligence, is examined in order to, starting from the current contribution of the neurosciences, especially with regard to the function of emotionality in the decision-making process, point out the new direction and scope of vulnerabilities in the information society and, in this sense, outline some guidelines for their regulation, beyond the protection of personal data, and, therefore, a conformation with the multilevel protection of the human person inside and outside the virtual/digital ecosystem. The scientific method involves a deductive approach, with explanatory, exploratory research and bibliographic procedure.

KEYWORDS: Artificial intelligence; neurosciences; human and fundamental rights; information society; data protection.

1 Orcid: https://orcid.org/0000-0003-3628-0852.

SUMÁRIO: 1 Premissas básicas; 2 Uma contribuição das neurociências para a análise dos processos de tomada de decisão do ser humano; 3 IA e as possibilidades de regulação/regulamentação com foco voltado para a proteção multinível da pessoa humana; 3.1 Dimensões ontológicas e gnoseológicas acerca da IA; 3.2 O problema do emprego excessivo/irreflexivo da IA; 3.3 O cenário brasileiro em face da sociedade informacional; 3.4 Molduras éticas e jurídicas para o uso da inteligência artificial; Síntese conclusiva; Referências.

1 PREMISSAS BÁSICAS

A despeito do que se pressupunha até há pouco tempo, não há mais uma espécie de separação nítida entre os mundos real, digital e virtual; consequentemente, as dimensões do tempo e do espaço se tornaram realinháveis à medida do uso exponencial das novas tecnologias assentadas no fenômeno da comunicação/informação[2].

Logo, evidencia-se cada vez mais o valor dos dados na sociedade informacional, ou seja, trata-se de uma sociedade impactada e transformada pelo incremento do emprego das Tecnologias de Informação (TIC), que, em outras palavras, alterando o curso e o tracejo do marco civilizatório, gerou uma variação surpreendente no mercado de *commodities*, porquanto os dados pessoais passaram a ser mais valiosos do que o petróleo e demais combustíveis fósseis (Soprana, 2019).

A realidade atual, à vista disso, tem se tornado cada vez mais complexa, eclodindo daí uma significativa sensação de insegurança naqueles que se ocupam de compreendê-la, de vivenciá-la e, em razão disso, de buscar e de propor formas de regulamentá-la/regulá-la.

Apesar do incremento do uso das novas mídias, da comunicação desenfreada, bem como dos benefícios oriundos dele e das suas repercussões mais flagrantes, o que se pode verificar é uma expressão contínua de novos padrões comportamentais, sobretudo quando se tem em mente a relação do ser humano com as chamadas novas tecnologias (Massis, 2020). De fato, emerge uma plêiade de indagações que se projetam, inclusive, sobre alguns padrões tradicionalmente aceitos, como os afetos aos parâmetros civilizató-

2 Na primeira década do século XXI, o número de pessoas conectadas à internet passou de 350 milhões para 2 bilhões. Além disso, nesse mesmo período, o número de pessoas com celulares passou de 750 milhões para 5 bilhões. A expectativa para o ano de 2025 é de a maior parte da população mundial estar com acesso à informação instantânea, sendo que, se for mantido o ritmo de crescimento de pessoas conectadas à internet, ter-se-á, na mencionada data, 8 bilhões de pessoas *online*. (SCHMIDT, Eric; COHEN, Jared. *The new digital age*: reshaping the future of people, nations and business. London: John Murray, 2014. p. 15)

rios, sobretudo conferindo ressignificação à identidade, à personalidade, à sociabilidade e à alteridade.

Não seria demasiado advertir, nessa altura, que a Organização Mundial de Saúde (doravante OMS) já reconhece algumas formas de adicção relacionadas com o uso da tecnologia (Brasken, 2019), em especial quando se trata do uso das redes sociais e de *games*, sobretudo em razão dos impactos que podem ser causados à saúde psíquica dos usuários.

À guisa de ilustração, cada brasileiro gasta, em média, nove horas diárias conectado à internet, significando um somatório de 145 dias do ano conectados, particularmente por meio do uso de dispositivos celulares (Silva, 2019), que, por sua vez, adquirem cada vez mais o sentido de novas tornozeleiras eletrônicas[3] em uma ambiência eivada de controle, de sobrecarga de atenção e de vigilância (Han, 2020).

Essa tendência tem sido lugar comum em países em desenvolvimento, novas colônias do âmbito digital, e notadamente se se observa, *e.g.*, a quantidade de tempo em que os brasileiros passam nas redes sociais (Volpato, 2021). De fato, o ponto de partida, além de outros mecanismos de proteção, para a discussão sobre o uso das tecnologias na sociedade informacional pressupõe um sistema ativo, eficaz, seguro e factível de proteção de dados[4] que restaure, confira e promova a confiabilidade, a autenticidade, a integridade e a confidencialidade em termos que não se restringem à ideia e ao direito à privacidade.

Por outro lado, evidencia-se, cada vez mais, a questão referente à divisão digital (Almeida, 2005) em seus diversos níveis, que, em outras palavras, expressa, adensa e atualiza a desigualdade social como a face mais obscura e excludente de países como o Brasil (Gavras, 2020), abalando, portanto, de forma indelével, o exercício pleno da cidadania (Ceci, 2020).

Não se pode descuidar de que a relação do ser humano com as tecnologias de informação e de comunicação (doravante TICs), em síntese,

3 Cf. Medida Provisória nº 954, de 2020 (compartilhamento de dados por empresas de telecomunicações durante a emergência de saúde pública), alvo da ADIn 6387, que, em sede de cautelar, o STF declarou a existência de um direito fundamental autônomo à proteção de dados no Brasil. Disponível em: http://www.stf.jus.br/arquivo/cms/noticiaNoticiaStf/anexo/ADI6387MC.pdf. Acesso em: 2 jul. 2020.

4 A parte referente à ANPD – Autoridade Nacional de Proteção de Dados – foi vetada pelo então Presidente da República e, posteriormente, foi instituída por meio da MP que fora convertida na Lei nº 13.853/2019. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2019/lei/l13853.htm. Acesso em: 2 jul. 2020.

consiste em um ponto de inflexão em todas as áreas. Com efeito, desde a moral, perpassando pela ética, pela ciência, pelo Direito e pela economia, alterações profundas estão sendo efetivadas, perfectibilizadas e naturalizadas, particularmente quanto à percepção do tempo e do espaço, ou seja, no âmbito da subjetividade, da responsabilidade, da memória, da privacidade e, em especial, da autonomia (Vesce, 2020).

O ser humano parametrizado pelas TICs apresenta, em sua prefiguração, como uma esfinge tebana, algumas nuances comportamentais que, de fato, evocam parâmetros normativos inusitados, mas, que, de todo modo, implicam sempre em se revisitar as narrativas humanistas (Dobelli, 2019, p. 131-134; Serres, 2015, p. 265; Hawking, 2015, p. 226-229) já consolidadas, particularmente no que se refere à filosofia dos valores e à ética, alinhados ao catálogo de direitos já consolidados.

Portanto, na medida em que se desdobra entre os mundos real, digital e virtual (Rothblatt, 2016, p. 63-65), o ser humano, em seu novo perfil, aponta para a urgência em se estabelecer alguns padrões de regulamentação e de regulação (Doneda, 2020, p. 24) e, nesse caso, em emular algumas pautas de soluções para conflitos atuais e antigos que foram reinventados desde os surgidos em razão da exclusão social/digital e do uso intenso das chamadas novas tecnologias, mas, sobretudo, os que advêm a partir de uma minudente análise acerca da inteligência artificial (Hoffmann-Riem, 2019, p. 11-38) (doravante IA) e de suas aplicações (Weidenfeld; Nida-Rumelin, 2018, p. 53).

A propósito, torna-se inolvidável que a presença do ser humano na era informacional consiste em uma sistemática renúncia, inclusive, à condição tradicional de ser racional mediante a sujeição a uma ordem mundial controlada por um punhado[5] de empresas privadas que se utilizam dos dados pessoais (Doneda, 2020, p. 33), sobretudo os assim chamados sensíveis e, dentre eles, os biométricos (Sarlet; Caldeira, 2019, p. 1-27), em uma frenética algoritmização do cotidiano.

O condicionamento advindo da relação do ser humano com as novas tecnologias diz respeito, portanto, ao desejo intrínseco de ser incluído, de fazer parte, e que está atrelado ao funcionamento do sistema dopaminérgi-

5 Exemplifica-se com a atuação da Akamai no parâmetro global da proteção de dados.

co, pois se trata de uma sequência contínua de estímulos para a liberação da dopamina.

Há, por assim dizer, uma psicologia do condicionamento associada à economia da atenção e do vigilantismo que pode ser enfocada a partir dos faróis da função da emocionalidade, em geral, e, mais especificamente, mediante os recentes contributos das neurociências nessa área do saber para clarificar as possibilidades de gatilhos/riscos na seara das novas tecnologias e que podem ser encarados na medida em que se aprofundam os estudos sobre a chamada frontalização.

Importa, logo, afirmar que a dependência não é em tecnologia, mas nos mecanismos que estão inseridos no funcionamento e no *design* de algumas plataformas e que, por isso, servem para disparar injeções de dopamina desencadeadas no cérebro do usuário, ativando excessivamente o sistema glutamatérgico.

Urge assinalar que essa investigação de natureza bibliográfica, multidisciplinar e exploratória, empregando o método hipotético-dedutivo, intenta, por meio da atual contribuição das neurociências, pautando-se, em parte, no atual sistema normativo brasileiro e, em especial, no que toca aos achados neurocientíficos e às atuais contribuições jurídicas na seara da regulação acerca dos processos decisórios, propor algumas possibilidades de regulamentação que reordenem o sentido das fronteiras ético-jurídicas à empregabilidade desenfreada/abusiva e irreflexiva do conjunto de tecnologias disruptivas e pervasivas, mas, em especial, as conhecidas como IA.

2 UMA CONTRIBUIÇÃO DAS NEUROCIÊNCIAS PARA A ANÁLISE DOS PROCESSOS DE TOMADA DE DECISÃO DO SER HUMANO

Estudo plural que, por meio da análise do funcionamento cerebral, oferece *insights* para áreas como a biologia, a economia, a ética, a filosofia, a psicologia, a política, o Direito, dentre outras (Lima *et al*., 2017): as neurociências, dito de outro modo, estão voltadas para decifrar a complexidade que caracteriza o funcionamento e a anatomia do cérebro, mas, em particular, do sistema nervoso central; assim, trata-se de um conhecimento que se encontra em contínua evolução na medida em que evolui a partir dos avanços advindos da tecnologia aplicada e que compreende áreas como, *e.g.*, a neurociência molecular, a neurociência cognitiva e a neurociência comportamental.

Dedica-se, inclusive, a entender o modo como se manifestam as lesões cerebrais, inicialmente em uma perspectiva reabilitacional, que, de qualquer sorte, podem afetar a grande parte da população e, atualmente, têm se pronunciado de modo mais significativo em razão de acidentes e de traumas, das altas taxas de cortisol em razão do estresse diário nos grandes centros urbanos e da longevidade dos idosos (Bernardi, 2019). O incremento das pesquisas em neurociências advém, consequentemente, da intensa necessidade de dar um sentido novo de prolongamento e de produtividade ao corpo, mas, igualmente, de clarificar o funcionamento cerebral e, dessa forma, acrescentar outros sentidos à ideia de qualidade à vida humana (Goleman, 2014, p. 16-17).

A frenologia, no século XIX, pode ser apontada como um dos momentos precursores desse conhecimento. A análise topográfica do cérebro passou, assim como o estudo das lesões cerebrais e do déficit funcional, a ser elemento nuclear para o que atualmente se torna mais nítido com a compreensão neuroanatômica, neuronal e o uso das neuroimagens, isto é, o uso das ressonâncias funcionais (Tavares *et al.*, 2019).

Dentre os diversos sistemas que compõem o ser humano, reconhece-se, mediante a abordagem das neurociências, a complexa atuação das emoções, que passaram a ser consideradas como a força motriz do ser humano. Entende-se, então, o sistema nervoso como um mosaico de regiões em que cada uma delas tem uma função e uma organização específica, conectando-se de modo complexo e interativo, em que não há função mental pura, existindo uma combinação de ações psicológicas e fisiológicas para cada ato (Herculano-Houzel, 2020, p. 349-375). O sistema nervoso age, sobretudo, em conjunto com o sistema endócrino (Lin *et al.*, 2020), para, em suma, atuar em duas frentes básicas, ou seja, na manutenção da homeostase e na geração dos comportamentos humanos (Martins, 2015, p. 36).

Assinala-se que o cérebro humano é uma síntese da própria evolução: organiza-se como um somatório funcional de estruturas mais rústicas e de regiões mais sofisticadas e, assim, mais recentemente integradas à composição anatômica do ser humano, como as regiões corticais, perfazendo uma orquestração fina, porém intrinsecamente tensionada (Marhounová *et al.*, 2019).

Uma das mais emblemáticas contribuições das neurociências, além de desnudar os padrões universais das emoções básicas que do estudo do cérebro põe abaixo qualquer justificativa plausível para a discriminação

(Mendes; Mattiuzzo, 2019, p. 39) entre as pessoas e, elucidando a relação do ser humano consigo e com outrem, diz respeito ao resgate das ideias de cérebro social e, em especial, de inconsciente, que, mediante uma revisão radical, ganhou novos contornos, manifestando-se como uma espécie de momento anterior à consciência e que atua de modo determinante, reposicionando categorias como a autonomia e o livre arbítrio, que estão, em síntese, na base do sistema social e do processo civilizatório e que são afetadas pela superexposição às novas tecnologias.

O processo decisório, por sua vez, relaciona-se com o tempo, mais especificamente com o futuro, próximo ou distante e, nessa medida, com o processo de recompensa (*núcleo accumbens*), perpassando pela orquestração de um *pool* de estruturas que estão ligadas ao controle volitivo e à formação de hábitos. A bem da verdade, as condutas humanas são vetores resultantes entre a razão e a emoção, sendo que, na grande maioria, há uma modulação não necessariamente simples que se situa entre a dimensão emotiva e a racional. Dito isso, vale reafirmar que o problema central se volta sempre para as escolhas que se dividem no tocante às recompensas, ou seja, ao aspecto qualitativo, temporal e quantitativo, tocando necessariamente na ideia de uma espécie de frontalização que advém da relação com o fortalecimento das estruturas/instituições sociais e, assim, dos modelos/pautas de regulamentação e de regulação.

Vale reafirmar que o processo decisório envolve toda a complexidade anatômica do cérebro humano, vez que afeta diretamente as regiões frontais e os circuitos de navegação social, mas, de fato, consiste em um processo que afeta o entrelaçamento entre as áreas corticais e subcorticais. Nesse sentido, importa lembrar que a decisão implica, em diversos e distintos graus, o controle de emoções e de instintos, contextualizando-se a partir das memórias. Essa memória, impende sublinhar, por oportuno, consiste em ato essencialmente criativo.

Assim, na medida em que se fortalecem os vínculos e os aspectos/dimensões civilizacionais, o controle social se torna mais relevante, em particular no que se refere ao abafamento da emocionalidade, que, de modo geral, até recentemente, era tomada como um ponto negativo que se opunha à tomada de decisão racional. Contudo, tem se mostrado como um sistema ainda pouco explorado que pode oferecer grandes esperanças à potencialidade do cérebro em opor resistências ao que o vulnerabiliza e, deste modo, traduz-se em uma área de estudos neurocientíficos que contribui na geração de um incremento na capacidade adaptativa para os humanos.

Deve-se lembrar de que a responsabilização individual envolve as funções executivas e os processos neurais decisórios, ou seja, a decisão será sempre uma síntese temporal em que o passado, o presente e o futuro estão radicalmente relacionados. Assim, a decisão sempre se baseia no passado e na projeção do futuro.

O processo decisório se inicia a partir de estímulos, internos e externos, que, em síntese, são prontamente atendidos, mas que dependem de variáveis, vez que há uma relação entre o estímulo, a resposta e as consequências, levando em consideração as funções neuropsicológicas/executivas. Nesse sentido, deve-se apontar para a relevância da inibição ou do controle de impulso, das memórias operacionais que formatam a capacidade de dar significação às memórias que são relevantes, trazendo-as ao campo atencional, além da capacidade de antecipação de cenários futuros de médio e de longo prazo.

O sistema emocional, portanto, relaciona-se a uma vantagem evolutiva na medida em que serve na marcação das experiências com valência positiva e negativa, não havendo, com efeito, emoção boa ou má, tratando-se de indicadores de itinerários de vida que devem servir para incremento da adaptabilidade ao ecossistema. Trata-se de um sistema precedente ao emprego da racionalidade tradicional e, em razão disso, pode ser utilizado como um marcador da universalidade, vez que há o que se chama de emoções básicas em todos os grupos humanos. Dentre elas, *e.g.*, encontra-se o medo que está atrelado ao funcionamento da amígdala.

Outro ponto essencial diz respeito ao fato de que se trata de um sistema que, tal quais os mecanismos que embasam a utilização da IA, age por perfilhamento/taxonomia, ou seja, marca eventos, cria tendências, reconhece padrões e constrói memórias que são, em suma, discursividades como as que subjazem na estrutura dos algoritmos. Ressalte-se que as estruturas relacionadas com os comportamentos instintivos são mais arcaicas, mais primitivas. Assim, a consistência biológica é prevalente em relação à força que atua na modulação comportamental, pois se encontra em processo continuado de fortalecimento/amadurecimento, inclusive mediante a educação, e tem necessariamente uma natureza que envolve a perspectiva da epigenética.

Dito de outro modo, o desenvolvimento cerebral é sempre crucial no processo de tomada de decisão na medida em que envolve vários paradoxos e, dessa forma, envolve o contexto que possui uma natureza cambiante

e as características da individualidade do sujeito. Além disso, destaque-se que os padrões normativos, sociais, éticos e jurídicos são parâmetros para circunscrever as opções de condutas e atuam na medida em que apontam para a adequação e a inadequação, ou seja, indicam e reforçam processos de recompensa mais duradouros em detrimento dos impulsos imediatistas.

Válido mencionar ainda que as patologias (psicopatias/sociopatias, que são os desvios da normalidade) consistem em conjuntos pluridimensionais que envolvem estruturas díspares que apostam principalmente nas recompensas imediatas. A ideia de uma vontade livremente arbitrada, destarte, envolve definitivamente algumas situações de sequestro efetuadas pela atuação amígdala referente ao medo, vez que o sistema cerebral vai, de qualquer modo, testando experiências e, assim, apontando e gerando desejos e motivações. Enfim, a formação de hábitos ou de memórias procedurais tem uma essência ligada à repetição de padrões de ação e ao prazer, isto é, às respostas recompensadoras.

Mudar hábitos, com isto, ocorre na medida em que há a extinção de uma espécie de memória procedural enquanto se introduzem novos estímulos repetitivos compensadores e competitivos que estão relacionados com a neuroquímica da vontade e, assim, atrelados à motivação para a mudança de condutas desadaptativas. O processo civilizatório, daí, encontra-se imbricado com o fortalecimento das estruturas corticais; logo, o seu adensamento ocorre em virtude da ativação apropriada do lóbulo frontal.

3 IA E AS POSSIBILIDADES DE REGULAÇÃO/REGULAMENTAÇÃO COM FOCO VOLTADO PARA A PROTEÇÃO MULTINÍVEL DA PESSOA HUMANA

IA produz padrões por se ocupar das inúmeras formas de perfilhamento, encontrando-se mais alinhada à emocionalidade do que à racionalidade na sua faceta mais tradicional (Henning, 2019, p. 163). Importa, nessa conjuntura, reafirmar que o ser humano, em sua perspectiva emocional, não se configura apenas como um ser eivado de fragilidades, sendo, de fato, um notório sobrevivente e, nesse sentido, um vencedor na luta das espécies, vez que a emoção pode e deve ser usada a seu favor.

De fato, há uma relação direta entre o funcionamento do cérebro e as condutas desencadeadas a partir de estímulos que demonstra uma tensão contínua entre o ser humano e o meio no qual ele se encontra que, em suma, tem o intuito de incrementar as condições de adaptabilidade em marcações de valências positivas e negativas as experiências vividas, tendo em

vista a ideia de recompensa em um equacionamento contínuo que envolve a dor e o prazer.

A investigação acerca da IA e das possibilidades de sua regulação, portanto, encontra-se profundamente atrelada ao adensamento na concreta percepção da circuitaria emocional, em particular das funções atribuídas às reações advindas a partir do funcionamento da amígdala (tendo especial atenção ao estado de estresse atualmente incrementado pela massiva exposição às novas tecnologias e agudizado nos tempos pandêmicos), vez que, além de alterar profundamente a vida, oferecem outros padrões de intelegibilidade e afetam às capacidades cognitivas, podendo servir de base para ações que podem culminar ora na emancipação, ora na subjugação da pessoa humana (Rodotá, 2008, p. 141).

Com efeito, as linhas de código que perfazem as IAs carecem de nuances, de subjetividade e de autocrítica. Não possuem, em seu atual estado, a capacidade para encetar juízos de valor e, nessa medida, agir com intencionalidade própria.

3.1 Dimensões ontológicas e gnoseológicas acerca da IA

Inteligência, até meados do século passado, era considerada um atributo humano, vez que consiste na capacidade de, utilizando o substrato biológico, produzir um raciocínio lógico pautado essencialmente na racionalidade e que estabelece, envolve e determina os processos de tomada de decisão. Não custa relembrar que o reconhecimento de um padrão inteligente era proporcionalmente distanciado das áreas intuitivas/emocionais do cérebro.

A inteligência, interessa grifar, aponta para processos cognitivos que tomam os dados como matéria-prima para a produção das diversas espécies de informação, que, por sua vez, se manifesta na forma de conhecimento, fazendo parte do cotidiano da Humanidade desde tempos mais remotos e que, em síntese, tem forjado juntamente com outras dimensões dos saberes o que se convencionou chamar de traço civilizatório.

A inteligência, portanto, parametriza-se em relação ao Humano, sobretudo tendo em vista as circunstâncias de vida de cada pessoa, pois tem dela, em suma, uma percepção eminente conjuntural. Inteligência é, nessa altura, um atributo humano, em um primeiro momento, mas, em face de uma atual modelagem advinda com o incremento científico do século XX

que contemplou novas alternativas, artificializadas, tonaliza, de modo complexo, a relação Humano-máquina.

Inteligência, sendo assim, deixa de ser tomada como um atributo exclusivamente humano para servir na caracterização de alguns artefatos e de máquinas que, em razão da evolução da ciência de dados, passaram a ser autorreferenciáveis. Alerta-se para o fato de se tratar de um novo formato em que a inteligência alcança patamares além dos convencionais e implica novas técnicas de aprendizagem eminentemente sutis, pervasivas e disruptivas.

A inteligência artificial, deve-se reconhecer, consiste em uma criação algorítmica destinada a cumprir finalidades determinadas e especificadas com base no recebimento de dados objetivos e estruturados para gerar resultados igualmente objetivos. Ainda há de se fazer menção aos *learners*, ou seja, aos algoritmos inteligentes que criam outros algoritmos.

Essa visão artificializada, todavia, impende uma funcionalização, ou seja, trata-se de uma opção de abordagem na qual o ser humano deve necessariamente preceder à tecnologia. Afirma-se, com isto, que a inteligência artificial, na medida em que consiste em uma espécie de tecnologia, deve estar a serviço do ser humano para, em sua atuação, auxiliá-lo no desafio emancipatório de viver como o principal protagonista no momento atual e no futuro e, então, deve estar alinhada ao fortalecimento de uma circuitaria emocional que favorece uma vida mais livre, responsável, solidária e autônoma, apesar do atual contexto instável, incerto, volátil e complexo.

A artificialização da inteligência, deve-se sublinhar, tem como suporte o uso de máquinas que, mediante o armazenamento, o tratamento e o compartilhamento de dados, passam a encetar algumas ações de reconhecimento, de perfilhamento, dentre outras, que produzem processos de natureza decisória equiparáveis aos humanos.

Assim, a inteligência se refere, destarte, a uma capacidade de, por meio da tecnologia, se alcançar panoramas informacionais muito além dos outrora conhecidos e, em decorrência disso, introduz outros critérios de tomada de decisão outrora desconhecidos ou negligenciados em função da absoluta incapacidade do cérebro humano em sua conformação atual de alcançar padrões de armazenamento ou de velocidade comparáveis aos que tipicamente são atribuídos aos computadores de ultima geração, em especial os quânticos.

Oportuno lembrar a possível aproximação do modo operante das novas tecnologias com o do sistema volitivo no cérebro humano, chegando à razoável hipótese de vir a suplantá-lo em algumas atividades. Inteligência artificial (IA), importante advertir, é um ramo da ciência da computação que se propõe a elaborar dispositivos que emulem a capacidade humana de raciocinar, de perceber, de tomar decisões e de resolver problemas. Consiste, sem dúvidas, em uma série de conjugações de natureza empírica que foram sendo testadas em distintos ecossistemas nos últimos anos e que, à vista disso, têm uma relação direta com o período do pós-guerra e uma incontestável sintonia fina com as transformações do mercado.

Inteligência artificial (IA) se propõe a manejar dispositivos que simulem a capacidade humana de raciocinar, de perceber, de tomar decisões e de resolver problemas, enfim, a capacidade de ser inteligente. Deve-se salientar que o elemento básico para uma caracterização da inteligência artificial encontra-se na dimensão do aprendizado e, então, está situado na formação de perfis taxinômicos que, dito de outro modo, baseiam-se em uma primeira etapa na produção de grandes análises a partir de grandes bancos de dados, orientando-se no presente momento cada vez mais para a granulagem.

Aponta-se, nessa altura, que, assim como o cérebro humano se reorganiza no processo de aprendizagem, há uma espécie de reorganização algorítmica subjacente quando se aprecia a relação chamada de IoT (Internet das Coisas) e igualmente no que concerne ao *machine learning* (aprendizado de máquina). Nesse sentido, interessa reafirmar que as técnicas de IA, em regra, mimetizam o funcionamento cerebral. Na aprendizagem por reforço, à guisa de exemplo, um sistema de IA aprende a otimizar a função de recompensa, reforçando-a de forma a aumentar a probabilidade de recorrência.

Em síntese, as tecnologias cognitivas se referem aos sistemas inteligentes capazes, por seu turno, de aprender e de tomar decisões não estruturadas e não programadas previamente. Na prática, a atuação algorítmica ocorre a partir de cálculos probabilísticos, resultando da multiplicação de um vetor de entrada com milhões de parâmetros cujos valores foram engendrados mediante treinamento.

Com efeito, não se pode olvidar da extrema relevância dos *big data* (Sarlet; Molinaro, 2019) para a compreensão da IA, pois, em virtude do grande volume de dados aos quais se refere em sua essência primordial,

oportuniza uma empregabilidade que pode resultar em novas formas para o enfrentamento de situações ditas insolúveis como a fome, a precariedade e a miséria, mas pode igualmente acarretar novas molduras de vulnerabilização da pessoa humana, seja por se tratar de repaginações de violações de direitos e de garantias já consagrados, seja pela introdução de novos modelos discriminatórios (Mendes; Mattiuzzo, 2019, p. 40) e excludentes.

Dentre os diversos desafios advindos com o aumento exponencial dos usos de IA que singularizou o século XXI, distingue-se que, em se tratando de uma multiplicidade de tecnologias, *v.g.*, sobressai o problema que toca nos limites éticos e jurídicos da utilização das máquinas autônomas. Ainda merece grifo o fato de que, do volume dos dados produzidos e em franca expansão, apenas um baixo percentual se encontra estruturado e, portanto, é, de fato, utilizado. Assim, há um amplo campo no que toca ao emprego de IA que se desdobra de forma contínua, generalizada e sem precedentes.

Em razão do uso cada vez mais adensado de IA no cotidiano, algumas máquinas, *e.g.*, os carros autônomos, passaram a ser fonte de questionamentos essencialmente voltados para o estabelecimento de limites de sua responsabilização e, deste modo, acerca do reconhecimento de novas formas de personalidade aplicáveis, de modo geral, aos robôs, aos computadores e, especificamente, aos algoritmos.

Outro elemento primordial que tem sido colocado no centro do debate se refere ao uso preditivo da internet que requer o autoconhecimento/a autopercepção como uma espécie de baliza para que o ser humano possa impedir a supremacia da máquina e, consequentemente, a chamada "Ditadura de Dados". Em certa medida, tornando-o apto a opor anteparos à algoritmização da vida. Dentre alguns aspectos nocivos, afirma-se a neutralidade apriorística e a recorrente ausência de explicabilidade e de justificação.

Trata-se, de fato, de um cenário atual em que há uma nova roupagem para o conceito de Humanidade, sobretudo mediante a imposição de uma configuração relacionada com uma projeção/modelo dos monopólios que atualmente regem a área da tecnologia e que necessariamente carecem de uma urgente transmutação na qual a opacidade deve ceder espaço à transparência, à auditabilidade e à compreensibilidade.

Importa reafirmar que, em rigor, o que se observa é que a IA funciona a partir da dinâmica de produção/criação/programação de algoritmos

que, discursivamente, implementam formas de resolução de problemas e, oportunamente, têm instaurado novos parâmetros para a cognição e para a decisão, ora incluindo, ora excluindo o agente humano. E, de qualquer sorte, não se pode olvidar que a natureza nuclear dos algoritmos de IA ainda é entendida como a de agente. Cognitivo, destarte, é uma expressão relacionada com o processo de aquisição de conhecimento. O conhecimento pode envolver diversos fatores, como o pensamento, a linguagem, a percepção, a memória e o raciocínio. Lembrando que a memória é eminentemente um ato emocional, criativo.

Atualmente, destacam-se as várias aplicações no cotidiano da inteligência artificial, *v.g.*, jogos, carros autônomos, programas de computador, aplicativos de segurança para sistemas informacionais, robótica, dispositivos para reconhecimentos da face, da íris, da escrita a mão e o reconhecimento de voz, programas de diagnósticos médicos, gestão de tráfego, decisões administrativas, serviços automatizados de assistência de veículos, *smart home*, artefatos inteligentes e artefatos na área militar, *smart* medicamentos na indústria farmacêutica e muito mais. Outras modalidades podem ser igualmente apontadas, como Alexa da Amazon (Canaltech, 2019), API. ai do Google, Sirikit da Apple e os modelos utilizados nas bolsas de valores de alguns países, como os EUA, e, inclusive, no Banco Central brasileiro. Igualmente se verifica o emprego de IA para potencializar silenciadores de armas, para o rastreio de crimes financeiros, para detectar tendências suicidas, para o controle da gestão pública e notadamente tem sido frequente em segurança pública.

Alguns pontos inquietantes sobejam, sobremaneira no que toca à responsabilização dos entes autônomos; aos limites dos algoritmos na tomada de decisão moral e jurídica; aos graus de afetação cerebral, à forma de jurisdição aplicável em casos de apuração de danos causados por IA; às implicações em razão da privacidade (Nissenbaum, 2010, p. 11) e dos direitos referenciados e emoldurados pela proteção de dados pessoais; à urgência no reconhecimento e na afirmação de patamares normativos extraterritoriais que impliquem cada vez mais práticas de colaboração de escala global. Enfim, há diversos pontos inquietantes e ainda em aberto que circundam a ideia de uma perfectibilização algorítmica em padrões democráticos e alinhados aos de segurança, de confiabilidade, de justiça, de liberdade, de dignidade e de cidadania.

3.2 O problema do emprego excessivo/irreflexivo da IA

A sociedade informacional, deve-se sublinhar, eminentemente pautada pela formação de redes controladas por algoritmos e alimentadas pelo fluxo contínuo de dados, não conhece limites nos moldes outrora validados na democracia liberal (Todorov, 2012, p. 197), vez que consiste em um fenômeno de caráter transfronteiriço, produzindo efeitos transgeracionais e que se expressa primordialmente mediante atos políticos que afetam radicalmente a ideia de soberania estatal. As regras estabelecidas pelo mundo digital, em certa medida, passaram a ser as balizas comportamentais dos seres humanos, dentro e fora dele, forjando um ambiente de perplexidades e de incertezas.

A IA, enfatiza-se, consiste em uma tecnologia essencialmente sutil, pervasiva e disruptiva que tem impacto ubíquo, podendo ser empregada em vários domínios, independentemente de plataforma. Dentre as inúmeras implicações, destaque-se a perspectiva ou, em outros termos, os perigos da substituição total do elemento humano e, de forma adensada, da singularidade que tem servido para o assombro de inúmeros cientistas e das pessoas em geral.

A singularidade não é uma ideia recente, remontando aos anos 70 do século passado, ocasião em que se tornou recorrente a questão acerca da superinteligência de máquinas. Apesar dos chamados invernos na história da tecnologia, o conceito de singularidade foi sendo reforçado na medida em que se alcançava patamares mais afinados entre o cotidiano e a produção de algoritmos superpotentes no contexto marcado pelo poder computacional que apontava em escala exponencial para a era das expressões das entidades não humanas como algo factível. Não custa dizer que o paradigma da singularidade vem perdendo sua força, ao passo que outros riscos são descortinados enquanto novas janelas de oportunidade para o ser humano se abrem.

Assim, divididos em entusiastas e opositores, os grandes estudiosos de tecnologia testemunharam a emergência de uma linha de produção sofisticada nessa área, sendo particularizada pela miniaturização, pela invulgar capacidade de aprendizado, pela opacidade e pelo alto grau de aplicabilidade prática, sobretudo no mercado.

Tendências apontam para o emprego de nanotecnologias para a produção de proteínas, para a expansão no emprego de IA na indústria

armamentista e para a manufatura de fármacos individualizados que, indubitavelmente, acarretarão em impactos sociais, econômicos e culturais desmedidos. IA se apresenta, de toda sorte, como uma saída/superação para as falhas humanas que, em regra, são atribuídas à excessiva emocionalidade que, até recentemente, havia sido considerada apenas como algo negativo.

Resta, contudo, um apelo para o aprofundamento no viés antropocentrista que, na medida em que demonstra os vácuos quanto ao conhecimento/mapeamento da totalidade do funcionamento cerebral e, com isto, aos limites da relação entre a inteligência humana e o sistema de emoções, esboça contornos em um pontilhado que tende ao infinito, ou seja, o ser humano aponta para infinitas possibilidades de superação e de realinhamento/resiliência cerebral. A inteligência artificial, nessa abordagem, poderia ser mediada pelos avanços no que toca às descobertas em relação ao ser humano e, nessa toada, ao funcionamento cerebral, em uma projeção que alcança suas inúmeras potencialidades, inclusive no que concerne à solução de *hard cases*.

Em rigor, o que não se pode olvidar é que o rol de condutas em um ecossistema balizado pelo binômio *Homem-máquina* envolve a rígida parametrização por meio da responsabilidade, da solidariedade para o devido gozo da liberdade, da dignidade e da autonomia, dentre outros direitos, especialmente o direito ao livre desenvolvimento da personalidade. Esse enquadramento que envolve o ser humano em relação às novas tecnologias pode se projetar e apontar para outro modo de utilização e, especialmente, de correlação/cooperação e de regulação/regulamentação.

De fato, o uso desenfreado da IA vai além dos limites da privacidade, podendo afetar paradigmas centrais como a isonomia, a igualdade, a democracia, a paleta de liberdades, afetando igualmente as capacidades cognitivas, os níveis de atenção, a autorreferenciação do Humano, dentre outros. E, por decorrência, pode acirrar a sociedade da vigilância (Bruno, 2013, p. 145), mediante uma algoritmização ilimitada, como partida necessária, mas não um fim em si mesma. Outras afetações podem ser elencadas no âmbito jurídico, em especial no que concerne ao direito das comunicações, ao direito de concorrência, à proteção do patrimônio intelectual, à esfera da responsabilidade civil, sobretudo no setor da responsabilidade por produtos, alterando, inclusive, as regras do mercado financeiro e da trabalhabilidade (Limberger, 2016).

À guisa de ilustração, algumas das chamadas novas tecnologias foram rechaçadas pela própria comunidade científica em virtude do desvelo da gama de prejuízos advindos com sua utilização, como o *Google Glass*, as técnicas *crisp babys*, os *e-cigarretes* e as cápsulas de café juntamente com as técnicas de reconhecimento facial (Silva, 2019). Há de se repensar, nessa toada, a necessidade de se pôr a crivo a ciência e de estabelecer fronteiras de atuação que restaurem a centralidade da pessoa humana em todos os sentidos, independentemente dos incentivos à inovação como base elementar para o desenvolvimento sustentável (Nações Unidas, 2020).

Oportuno é relembrar que a base da chamada revolução tecnológica cujo foco é projetado em particular pelo emprego da IA vem sendo encetada pela ação dos chamados *big five* e da atuação da China. Portanto, na arena global, passou a ser desenhado um panorama de monopólio em que os *top five* (Sundar Pichair, Apple, Microsoft, Amazon e Facebook) da tecnologia passaram a ter mais poder concentrado do que os Estados e, consequentemente, implicam uma nova leitura da soberania estatal, da territorialidade e do sentido de aplicabilidade do catálogo de direitos humanos e fundamentais (Marsden, 2010, p. 36). Vale referenciar que, nessa conjuntura geopolítica, tanto a União Europeia quanto a China têm se projetado como importantes atores na construção de sistemas normativos globais de proteção em face dos algoritmos.

A despeito desse movimento, a conduta dos agentes das chamadas *big five* da tecnologia ainda sinaliza para um extremo reducionismo em relação ao mundo digital/virtual no qual elas atuam na forma de grandes monopólios no sentido de que ainda não se reconhece uma pauta prioritária em relação à economia da atenção que sustenta o capitalismo cognitivo, que, por sua vez, encontra-se umbilicalmente atrelada ao estado de vigilantismo cada vez mais exacerbado[6] e sutil.

Desta feita, o influxo da tecnologia no cotidiano, além de concretizar alguns benefícios, fez eclodir novas formas de vulnerabilizações da pessoa humana, evocando, daí, a tarefa do jurista de participar do diálogo emulado

6 Conforme Assange (2015, p. 16): "O alerta sobre o Google é também um aviso acerca da natureza ambivalente das tecnologias de informação e comunicação. É a lembrança de que o poder não se faz por meio da tecnologia somente, mas está embutido na própria tecnologia. Redes digitais e seus dispositivos não são neutros. Seus arranjos e limites embarcados em protocolos e códigos são programados para cumprir determinações, muitas vezes de ordem geoestratégica, política e econômica. Um algoritmo do Google ou do Facebook funciona de um determinado modo não porque não haveria outra forma de funcionar, mas porque foi concebido daquele modo".

por inúmeras frentes de resistência ao mau emprego/desvio da tecnologia, que se expressa, sobretudo, em relação às camadas mais desfavorecidas da população. Cabe, em especial ao jurista, emular esforços para emitir outras nuances e constelações de direitos engendradas mediante emprego do sistema normativo em vigor que, nessa medida, se tornem factíveis no momento atual, particularmente com o intuito de assegurar e de salvaguardar mediante uma ressignificação dos patamares historicamente consagrados de direitos que, de certa forma, se encontram em plena erosão[7].

E, nesse sentido, oportuno alentar para a necessidade de combate e de enfrentamento das modalidades de injustiça algorítmica que vêm sendo alvo de denúncias e que ainda escapam de um olhar mais arguto por parte do Judiciário.

3.3 O cenário brasileiro em face da sociedade informacional

Segundo os relatórios do IPEA (2019) e do IBGE (2020), a questão se adensa quando a mira se volta para o Brasil, vez que se reconhece um quadro de especificidades nacionais como o fato de ser uma sociedade hiperconectada, uma das maiores consumidoras de tecnologia, um dos piores sinais de internet, um abismo composto pelas lacunas no ordenamento jurídico, um déficit educacional que afeta a formação de recursos humanos para atuar nessa área, um alto índice de corrupção, e de conivência dos agentes públicos com relação aos abusos cometidos pelas empresas de tecnologia e uma realidade marcadamente assimétrica que se tornou um campo fecundo para os efeitos da divisão digital que grassa nos dias atuais.

O índice de acesso digital (IAD) mede a capacidade global das pessoas, levando em consideração a qualidade, a infraestrutura, o conhecimento, a acessibilidade e a forma de utilização que perfazem o perfil de utilização de um Estado, consequentemente podendo ser utilizado para medir o grau de vulnerabilidade de uma população face às TICs (Tokarnia, 2020).

A propósito, a desigualdade no acesso à internet e às TICs se chama exclusão/divisão digital e afeta 52% das mulheres e 42% dos homens no mundo. Segundo os dados colhidos no portal Internet World Stats em 2020,

7 Conforme Leite (2016, p. 150): "Sociedade da Informação – que nada mais é do que uma forma específica de organização social em que a gestão, o processamento e a transmissão de informações tornam-se as fontes fundamentais de produção e de poder, devido às novas condições tecnológicas surgidas nesse período histórico. O surgimento dessa nova sociedade trouxe, portanto, a necessidade de repensar o papel do Estado nesse novo contexto".

na África 39,3% da população vive conectada a despeito dos 94,6% dos norte-americanos (IDC, 2021). Infere-se disso um abismo tecnológico desproporcional que separa alguns países e pessoas às vésperas da expansão do 5G e, nesse sentido, tende a aprofundar a desigualdade e a exclusão na medida em que não se pode mais negligenciar a intrínseca relação entre o acesso à internet e a educação digital, que, em suma, consiste no processo de aprendizagem que permite que uma pessoa adquira competências cognitivas para usufruir, de modo mais responsável e consciente, os potenciais educativos, econômicos e sociais das novas tecnologias.

Urge mencionar que o Brasil tem se transformado em um celeiro ou em uma espécie de fazenda digital em razão da ausência de um sistema protetivo robusto e eficaz e, inclusive, de investimentos de base nas áreas da educação, da ciência e da inovação que encetem a produção de uma série de canais voltados para fortalecimento dos pilares da cidadania digital, em particular voltados para a proteção dos grupos mais vulneráveis, que são as crianças, os adolescentes e os idosos (Scola, 2020).

Diante de um horizonte em que há inúmeras pessoas invisíveis ao Estado brasileiro portando um celular e, nessa mesma perspectiva, há um contingente de 29% da população de desbancarizados (Agência Brasil, 2019), a paisagem nacional se torna invulgar e exige providências para, partindo de uma análise lúcida, séria, comprometida e contínua, possam ser desvelados um conjunto de características próprias que implicam medidas emergenciais na produção de pautas de solução a médio e a longo prazo e que, com efeito, apontem para a concretização de mudanças estruturantes que capacitem a população para atuar na medida de um protagonismo responsável, autônomo, livre e solidário em face dos danos advindos com a exposição exagerada/despreparada às TICs.

Lembrando, nessa altura, que a exclusão digital ocasiona, dentre outros efeitos, a incomunicabilidade, o isolamento, a disparidade na formação educacional, diminuindo as possibilidades de qualitativo ingresso no mercado de trabalho e, principalmente, aprofunda a discriminação, tanto sexual quanto racial (Ribeiro, 2019, p. 43-44).

Reconhece-se que, embora a Constituição Federal de 1988 não seja propriamente digital, o Brasil já conta com um ordenamento ancorado em princípios e em dispositivos que garantem o gozo de direitos como a intimidade, a igualdade, a liberdade, a segurança, a não discriminação e o livre desenvolvimento da personalidade que, por si, já consolidam um esteio

significativo para se opor resistência a alguns dos desvios outrora mencionados.

Além disso, há sempre um alento ao se contemplar a efeméride dos trinta anos do Código de Defesa do Consumidor e do Estatuto da Criança e do Adolescente, bem como o corolário de direitos advindos com o Marco Civil da Internet, comumente chamado de Constituição da Internet no Brasil. E, ao relembrar a trajetória da consolidação do direito do consumidor no Brasil, emerge uma ideia um tanto alvissareira de que a dimensão da cidadania digital/virtual possa trilhar igualmente um caminho exitoso.

Impende sublinhar o teor da Lei nº 13.787/2018, que consiste em uma normatização voltada para a proteção dos dados dos pacientes nos prontuários eletrônicos. Essa legislação garante a preservação dos dados sensíveis dos pacientes. Para tanto, mecanismos como a restrição de acesso e a criptografia dos dados são pontos/chaves que se tornam capazes de garantir a privacidade dos pacientes e, de outra banda, a confidencialidade dos atos dos profissionais de saúde.

Nesse ensejo, destaca-se igualmente o movimento que gerou a produção legislativa da Lei Geral de Proteção de Dados (doravante LGPD) em 2018 e, nessa mesma ordem, que impulsionou posteriormente a propositura do PL da LGPD penal em tramitação no Congresso. Com efeito, o ordenamento jurídico brasileiro, em parte afetado pela necessidade do país de efetivar a participação na Organização para a Cooperação e Desenvolvimento Econômico (doravante OCDE), produziu uma lei tocada pelo padrão europeu (GPDR), restando ainda lacunas a serem preenchidas pelo Judiciário, pelo Parlamento, pela Autoridade Nacional de Proteção de Dados (doravante ANPD) e pelo Conselho Nacional de Proteção de Dados e Privacidade (doravante CNPD).

A Lei nº 13.709/2018 considera, em seu art. 2º, VII, como fundamento *os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais*. Esse fundamento, particularmente, está associado diretamente ao consentimento das pessoas naturais expresso como elemento transversal na lei ao longo de seus 65 artigos, tanto no que concerne a sua exigência, à renuncia, como a sua dispensa. Oportuno destacar que, a despeito do entendimento anterior tocado pela influência europeia de consistir em uma base legal superior, o consentimento ainda prefigura como um dos pilares da proteção de dados.

De mais a mais, é exigência da LGPD o consentimento informado para o tratamento de dados pessoais, normatizado no Capítulo II, art. 7º, I, conjugado com o art. 8º e seus parágrafos, que estabelecem a forma (§ 1º), as responsabilidades do controlador (§§ 2º e 6º), as vedações e os critérios de nulidade (§§ 3º, 4º e 5º). Por sua vez, o art. 9º determina, com base no princípio do livre acesso, que os direitos do titular dos dados pessoais podem ser empregados para obter informações sobre a finalidade, a forma, a identificação e informações do controlador, o compartilhamento realizado pelo controlador e suas respectivas responsabilidades, entre outras características, previstas do inciso I ao inciso VII, além dos direitos estabelecidos no art. 18.

As exceções previstas ao consentimento informado elencadas nos § 4º do art. 7º da LGPD – isto é, a dispensa *do consentimento para os dados tornados manifestamente públicos pelo titular*, resguardados os direitos do titular e os princípios legais, em especial a *finalidade, a boa-fé e o interesse público*, previstos no § 3º – não esclarecem o que é tornar manifestamente públicos dados e informações.

O Capítulo III da LGPD deve ser destacado, pois nele são previstos os direitos dos titulares, estabelecidos mais especificamente nos arts. 17 a 22. O artigo introdutório do Capítulo, art. 17, determina as bases inerentes ao pleno desenvolvimento da personalidade – isto é, considerando *os direitos fundamentais de liberdade, de intimidade e de privacidade* –, direitos que serão desdobrados e especificados *intra legis* ao longo da lei, especificamente nos artigos seguintes – arts. 18 a 22. Entretanto, esses direitos devem ser, também, sistematicamente interpretados *extra legis*, particularmente considerando a Constituição Federal, o Estatuto da Criança e do Adolescente (ECA) e os Códigos Civil e do Consumidor.

Cabe o destaque ao princípio do melhor interesse das crianças e dos adolescentes previsto na Seção III, art. 14, que exige, em seu § 1º, o consentimento de pelo menos um dos pais ou do representante legal e prevê limitações para a coleta, o armazenamento e o tratamento postos nos parágrafos seguintes[8].

8 Conforme Self *et. al.* (2017, p. 366-376): "*Assent is respectful of minor patients and ensures their rights are protected. Policies and procedures can be developed to incorporate assent, researchers have a stronger conceptual base, and educators can train nurses to help pediatric patients develop medical decision making skills*".

Um aspecto notável foi o fortalecimento da proteção e a decorrente vedação de uso ilegal e nocivo de dados sensíveis e, dentre eles, os biométricos, particularmente os dados referentes à saúde, para fins discriminatórios independentemente do consentimento do usuário, sobretudo em face dos riscos de destruição, de divulgação e de acesso indevido em razão da estrutura aberta da internet, previstos na Seção II, arts. 11, 12 e 13 da Lei nº 13.709/2018. E, nessa altura, urge evidenciar a incerteza quanto ao rol do qual se depreende uma falta de definição exata de dados sensíveis e que, por sua vez, implica um trabalho interpretativo e argumentativo posterior.

De qualquer sorte, extrai-se desse texto legislativo, além de relevantes conceituações como as que consagram a principiologia, que diferenciam os dados pessoais e a sua relevância no contexto atual a despeito da especificação exata em relação aos sensíveis, a fundamentalidade e o âmbito de proteção do direito fundamental à proteção de dados pessoais, recentemente afiançado em sede de controle de constitucionalidade pelo Supremo Tribunal Federal (STF) mediante julgamento histórico em que suspendeu a eficácia da Medida Provisória (MP) nº 954/2020. Importa salientar que, naquela sessão emblemática, foi engendrada uma mutação constitucional baseada na lógica de que não há dados irrelevantes, neutros ou insignificantes; assim, restou reafirmada a proteção constitucional ao dado pessoal e, nessa medida, à pessoa humana no ecossistema digital/virtual.

Interessa admitir que houve a afirmação de um direito fundamental autônomo e, na outra face, a afirmação de um duplo dever do Estado brasileiro, ou seja, de um lado, a tarefa de se abster de interferir negativamente no âmbito de proteção desse direito e, de outra banda, de adotar as medidas apropriadas e que assegurem o seu devido cumprimento e sua concretização apropriada.

Nesse ambiente ainda mais tensionado pela eclosão da situação pandêmica, agravado ainda em razão dos equívocos relacionados com a ausência de criação/implantação de uma Autoridade Nacional de Proteção de Dados (ANPD) livre de qualquer interferência do Poder Executivo, evidencia-se como lapidar o papel do constituinte derivado, da doutrina, da jurisprudência e, naturalmente, da vivacidade dos diálogos na academia para a produção de novos parâmetros, limites e fronteiras.

Para além desse panorama, deve ser mencionada a recente entrada em vigor da parte punitiva da LGPD, a instituição do Conselho Nacional de Proteção de Dados Pessoais e Privacidade, bem como as audiências públi-

cas e de uma série de medidas tomadas pela ANPD. O fato é que, já com base na principiologia constitucional, no catálogo de direitos e de garantias fundamentais, bem como dos direitos extraídos do ordenamento jurídico brasileiro, alinhavados em uma constelação normativa pelos direitos humanos, torna-se possível atuar positivamente nessa seara.

Apropriadamente deve ser mencionado o relatório de 2021 no teor da transparência digital que, dentre outros achados, concluiu que, em se tratando da esfera federal, há uma aceleração no emprego da IA, caracterizando-se por dificuldade quanto à observância de parâmetros éticos e jurídicos, automatismo na tomada de decisão e uma ideia de neutralidade, falta de capacitação técnica e ética dos agentes públicos, insuficiência de conhecimentos e reflexão acerca da acurácia da IA e insuficiência de discussão e de conscientização no que se refere às injustiças algorítmicas (IPEA, 2021).

O Governo Federal, mais especificamente o Ministério da Ciência, Tecnologia e Inovações, por sua vez, lançou as estratégias de IA, distinguindo três grandes eixos transversais: legislação, regulação e uso ético; governança de IA e aspectos internacionais; em confluência com os eixos verticais: educação, força de trabalho e capacitação, PD&I e empreendedorismo, aplicação nos setores produtivos, aplicação no Poder Público e em segurança pública (Brasil, 2021). A Estratégia Brasileira de Inteligência Artificial (EBIA) tem um papel de nortear as ações do Estado brasileiro em prol do desenvolvimento das ações, em suas várias vertentes, que estimulem a pesquisa, a inovação e o desenvolvimento de soluções em inteligência artificial, bem como seu uso consciente e ético, sendo sistematicamente alvo de críticas pela sua amplitude, vagueza e falta de pragmatismo, de objetividade e de uma prognose adequada aos seus propósitos, ao tamanho e à atual situação do Estado brasileiro.

Interessa lembrar que a proteção de dados é, como já se advertiu, o ponto elementar para a composição de um sistema protetivo adequado que prefigure condições de assegurar limites éticos para o emprego de IA – tanto na esfera pública quanto privada. Nessa sequência, alerta-se aqui para o teor do art. 20 da LGPD, que, embora de forma truncada, oportunizou e consagrou o direito à explicação e à revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais.

Além da entrada em vigor da LGPD em sua totalidade, que, em certa medida, deve ser entendida como um *step one*, o CNJ – Conselho Nacional de Justiça –, no que toca à proteção de dados públicos, editou resolução a

respeito da utilização da IA no âmbito do Poder Judiciário (CNJ, 2020). E, tal qual se mencionou, há um PL em tramitação da vulgarmente chamada de LGPD penal que se destina à proteção no âmbito dos dados sob a gestão do setor público.

Essa resolução do CNJ, *e.g.*, trata das vantagens como ganhos na agilidade na prestação jurisdicional, ressaltando, contudo, que as decisões judiciais apoiadas pela inteligência artificial devem preservar a igualdade, a não discriminação, a pluralidade, a solidariedade e o julgamento justo, ou seja, por meio da viabilização de meios destinados a eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos (Velasco *et al.*, 2020).

The last but not the least, devem ser apontados os projetos de lei em tramitação nas casas legislativas do Congresso Nacional sobre IA e, dentre eles, PL 5.051/2019 e PL 5.691/2019, em tramitação no Senado Federal, bem como os PL 21/2020 e PL 240/2020 da Câmara dos Deputados, importando, inclusive, citar a PEC (projeto de Emenda Constitucional nº 17).

3.4 Molduras éticas e jurídicas para o uso da inteligência artificial

Tecnologias como IA dotadas de potencial outrora mencionado e pautadas em uma nova configuração da inteligência implicam uma estruturação alicerçada nos seguintes princípios: transparência; auditabilidade; responsabilidade; imputabilidade; reversibilidade; proteção contra o uso de vieses discriminatórios, sobretudo aqueles ocultos pelos chamados algoritmos cuja opacidade os delineia como *black boxes*; proteção efetiva de dados pessoais, em particular no que toca aos dados pessoais sensíveis e em amplo consenso popular quanto à urgência por novos parâmetros para o consentimento livre, autêntico e genuíno.

Oportuno lembrar que, diante desse panorama, urge redimensionar a privacidade de modo que, enquanto direito humano e fundamental, venha a comportar outras acepções além da individual, ou seja, encetando uma perspectivação de contexto e, nesse sentido, uma dimensão coletiva e transindividual, em particular quando se volta a atenção para os agravos que podem macular o regime democrático como os que se infere com as revelações do escândalo envolvendo a *Cambridge analytical*.

Deve-se sublinhar a urgência por molduras/parâmetros que sejam adequados a tornar o panorama global mais seguro, confiável, integro,

autentico, sobretudo quando se volta a mirada para os países em desenvolvimento que, na atualidade, têm se notabilizado como grandes celeiros digitais, cuja população tem, em geral, sofrido uma intensa e continuada manipulação, direta e indireta, em sua estruturação identitária e democrática e, em decorrência, tem sido muito afetada em suas capacidades cognitivas, principalmente no que toca ao processo decisório e na adequada confluência de elementos racionais e volitivos para a devida frontalização. Trata-se de um reposicionamento que, inclusive, alinhe a diversidade e a ética à tecnologia.

Atualmente, o que se verifica, de fato, é uma tendência à adoção de parâmetros jurídicos em rede, bem como de um punhado de regras de *compliance* digital em nível global, além da adoção de regras algorítmicas que, respeitando os limites tênues do que toca ao segredo industrial e à propriedade intelectual, manifestem uma natureza apropriada à *accountability* e, nesse sentido, afetas à responsabilização e à transparência. Faz sentido ainda lembrar a clássica contribuição de Assimov.

As implicações futuras exigem novos *designs* de caráter político, jurídico e econômico, em especial quando se torna cada vez mais perceptível que a questão não se circunscreve meramente ao âmbito patrimonial, vez que, dentre outros aspectos, possui uma natureza multirrelacional. A atenção se volta igualmente para a qualidade dos dados e, dessa maneira, configura-se cada vez mais a complexidade do tema.

No plano internacional, ganham sentido as contribuições advindas da declaração de Montreal, dos princípios de Asilomar, da declaração de Toronto, da Comunicação da EU sobre inteligência artificial para a Europa e das diretrizes produzidas pela OCDE.

Interessante, nessa altura, mencionar o chamado livro branco sobre a IA, iniciativa europeia que, em síntese, propõe pilares para a excelência e para a governança nessa área. Destaque-se que a iniciativa visa preconizar uma atuação conjunta que concilie os interesses dos cidadãos/usuários em geral, os interesses empresariais e, assim, visa assegurar o interesse público de modo a promover o desenvolvimento e a inovação na medida em que se baseia na fundamentalidade de direitos e de garantias essenciais como a plêiade de liberdades, as faces formal e material da igualdade e, assim, a dignidade para o esboço da nova moldura da privacidade dentro e fora do ecossistema digital/virtual. Nessa altura, merece grifo o imenso desafio

enfrentado atualmente na União Europeia para esboçar padrões éticos para o emprego da IA, sobretudo na área da saúde[9].

Nesse entendimento, lança diretrizes que se inclinam para a excelência ao passo em que devem ser adensadas as parcerias público-privadas, o investimento na inovação e em políticas públicas de caráter estruturante que, acima de tudo, promovam a inclusão digital em seus diversos níveis e, assim, promovam a própria pessoa humana à posição preferencial (União Europeia, 2020).

Para tanto, apostam em grandes pontos como imprescindíveis à parametrização/regulação/regulamentação da IA, isto é, à garantia da iniciativa da criação e do controle final por seres humanos, à existência de um sistema robusto e seguro eminentemente voltado para o respeito, para a proteção e para a promoção da privacidade tanto quanto para uma política adequada de governança de dados, em especial atenção para com os dados sensíveis e, nessa perspectiva, alicerçado na equidade, na responsabilização, no bem-estar social e ambiental e na inclusão, resguardando à não discriminação como direito humano e fundamental.

Aqui se deve sublinhar que, segundo a posição da EU, a IA não deve ser entendida como um fim em si mesma, devendo ser funcionalizada, *v.g.*, à realização dos Objetivos de Desenvolvimento Sustentável da ONU, designadamente a promoção da igualdade de gênero, o combate às alterações climáticas, a racionalização da utilização de recursos naturais, a melhoria da saúde, da mobilidade e dos processos de produção, bem como apoiar a monitorização dos progressos alcançados com base em indicadores de sustentabilidade e coesão social.

Segundo a EU, devem ser enfatizadas as dimensões da solidez e da confiança, sendo a última entendida a partir de três componentes: a legalidade, a ética e a adequação segura tanto do ponto de vista técnico quanto do social. E, para isso, enfatiza a necessária expansão da ideia de responsabilidade para assumir as perspectivas individual e coletiva de tal sorte que possam evitar, inclusive, aspectos negativos não intencionais. Quanto à solidez, evidencia-se a adequação ao domínio da aplicação, ao ciclo de vida,

9 Tendências da área da saúde no que toca à telemedicina: crescimento das *healthtechs*, aceleração do processo de adoção de novas tecnologias na área da saúde, preocupação crescente com a segurança de dados na área da saúde, ampliação do uso de IA e de *compliance*, realização frequente de teleconsultas e de exames, bem como uso de prontuários eletrônicos.

além do elemento social que deve ser considerado no que toca ao contexto e ao ambiente em que o sistema deve operar. E, para tanto, verifica-se um papel central para os relatórios de impacto de riscos.

Significativo, em relação à atuação da EU no cenário internacional, é o teor da Proposta de Regulamento de IA (*Regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence*), que, dentre outros pontos, notabiliza-se por classificar os sistemas em três diferentes patamares de risco: inaceitável, elevado e mínimo. O desenvolvimento e a utilização de sistemas que apresentem risco inaceitável são vedados, enquanto os de alto risco são alvo de severas restrições no desenvolvimento, na implementação e no uso. No que se refere aos de baixo risco, as exigências dizem respeito à transparência. Há inovações no tocante à aplicação de regras e de obrigações que, em sentido alargado, aplicam-se a todos os que compõem a cadeia de produção de IA, caracterizando-se por ter um alcance extraterritorial (União Europeia, 2021).

Evidenciam-se, assim, três grandes tendências nessa temática: regulação por arquitetura, regulação a partir de paradigmas éticos e por meio de parâmetros jurídicos.

SÍNTESE CONCLUSIVA

Descartes sentenciou: "Penso, logo existo". Atualmente, diz-se que a capacidade cognitiva de ponderar e de decidir se encontra em uma espécie de derrocada no panteão das características humanas, particularmente quando se tem em mente o nível de alienação e, portanto, de renúncia e de comprometimento da racionalidade e da cognição em face das novas tecnologias e seus cantos de sirene.

Constata-se, de outra banda, a relevante posição da emocionalidade nos estudos e em diversas contribuições advindas das diversas áreas das neurociências para uma melhor percepção do agir humano e, por assim dizer, do funcionamento cerebral como um todo.

Obviamente que não se tem a pretensão de reduzir o ser humano ao funcionamento do próprio cérebro, mas trata-se de, por oportuno, alinhavar outras discursividades ao cenário em que se observa um crescente desassossego em relação às condições de vulnerabilização que se constata no incremento do uso das novas tecnologias e, em síntese, replicam e adensam a desigualdade, a estigmatização e os abismos de desigualdade do real no plano virtual.

Este é, sem sombra de dúvida, um horizonte arriscado que se delineia a partir do momento em que se cogita a substituição do Homem pela máquina, gerando a necessidade de se esboçar outros contornos para a cooperação entre eles. De fato, o que se observa nesses tempos pandêmicos é uma aceleração do emprego de IA e, de outra banda, uma maior irreflexibilidade quanto ao uso e quanto às consequências, seja em curto, médio ou longo prazo. Nessa altura, folga mencionar Dostoiésvski, que, em *Crime e castigo*, afirmou: "A inteligência por si só não é suficiente quando se trata de agir com sabedoria".

Assim, a percepção dos limites de empregabilidade da IA orienta-se para um panorama clarificado quanto à atual tessitura da condição de sujeito de direito, tendo em vista as inovações positivas, mas, consequentemente, atentando para os desafios e os riscos advindos quanto ao controle e à vigilância que podem se perenizar e até se agudizar na medida em que, *e.g.*, flexibiliza-se na concretização da proteção de dados enquanto direito humano primordial. Aqui se deve esclarecer que, em razão da necessidade de esforços no âmbito internacional, deve se tornar a percepção/concretização desse direito, projetando-se, na medida do possível e do razoável, para o que se infere da ideia de soberania de dados.

Em síntese, urge uma releitura do catálogo dos direitos humanos e fundamentais já consolidados na maioria dos países ocidentais para empreender esforços globais para a estruturação de pautas de governança que estejam apropriadas ao atual desenvolvimento de programas e de algoritmos e que sejam factíveis quanto à sua aplicabilidade, transparência e auditabilidade, garantindo a proteção multinível da pessoa humana e, com isto, garantindo a consolidação dos regimes democráticos alinhados com uma perspectiva de cibersegurança. Para tanto, entende-se a imprescindibilidade do contributo advindo das neurociências, sobretudo no que concerne ao processo de apropriação de outras dimensões comportamentais que podem ser fortalecidas para resultar em maior grau de emancipação da pessoa humana.

Assim, a investigação acerca dos trajetos neurocientíficos voltados para o estudo da emocionalidade, para tanto, sobretudo no que toca aos processos decisórios, passa a ser um caminho factível e imprescindível para a reordenação da ideia de inteligência de modo que a relação Homem-máquina possa se situar em uma harmonia ancorada em uma tessitura forjada por padrões éticos que reforçam o elemento humano como primordial. Para além disso, não se pode descurar das possibilidades advindas com a

necessidade de construção de sistemas de IA amigáveis à privacidade e, consequentemente, adequados à não interferência ilegítima e nociva na modulação de comportamentos.

A propósito, a inteligência artificial deve estar a serviço do ser humano para, em sua atuação, auxiliá-lo no desafio emancipatório de viver como o principal protagonista no momento atual e no futuro e, então, deve estar alinhada ao fortalecimento de uma circuitaria emocional que favorece uma vida mais livre, responsável, solidária e autônoma a despeito do atual contexto instável, incerto, volátil e complexo. Nesse sentido, salienta-se a particular significação dos relatórios de impacto de riscos.

Alguns pontos, todavia, permanecem em aberto, devendo ser considerados como alavancas arquimedianas, como: a possibilidade de integração dos atuais contornos jurídicos de responsabilização e de personalização no que se refere aos sistemas de IA; a necessidade de criação de padrões globais para a parametrização da IA; a delimitação do alcance moral e ético dos algoritmos; a definição de jurisdição internacional aplicável aos agentes autônomos.

A experiência europeia, dessa forma, pode e deve servir de base na configuração de uma nova realidade dentro e fora do mundo digital, na qual os demais países sejam chamados a assumir uma posição mais proativa no cenário mundial.

Quanto ao cenário brasileiro, urge uma ampla discussão pública que se projete na construção de um sistema factível e, consequentemente, mais apropriado à realidade que se projeta como inevitável, sobretudo na medida em que se avizinha a implantação do 5G e na medida em que o desenvolvimento e a utilização dos sistemas de IA devam assegurar transparência, auditabilidade, responsabilidade, explicabilidade, imputabilidade, reversibilidade, proteção contra o uso de vieses discriminatórios, proteção de dados pessoais e, de fato, parâmetros factíveis para o consentimento genuíno.

REFERÊNCIAS

AGÊNCIA BRASIL. Brasil tem 45 milhões de desbancarizados, diz pesquisa. Disponível em: https://epocanegocios.globo.com/Brasil/noticia/2019/08/brasil-tem-45-milhoes-de-desbancarizados-diz-pesquisa.html. Acesso em: 21 ago. 2020.

ALMEIDA, Lília Bilati de et al. O retrato da exclusão digital na sociedade brasileira. *Jistem J. Inf. Syst. Technol. Manag. (Online)*, São Paulo, v. 2, n. 1, p. 55-67, 2005.

Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1807-17752005000100005&lng=en&nrm=iso. Acesso em: 8 jun. 2020.

ASSANGE, Julian. *Wikileaks*: quando o Google encontrou o wikileaks. Trad. Cristine Yamagami. São Paulo: Boitempo Editorial, 2015.

BERNARDI, Jaqueline. Saúde da memória de idosos: projeto de intervenção. Disponível em: https://ares.unasus.gov.br/acervo/handle/ARES/12697. Acesso em: 8 jun. 2020.

BRASIL. Portaria GM nº 4.617, de 6 de abril de 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_portaria_mcti_4-617_2021.pdf. Acesso em: 24 set. 2021.

_____. Portaria MCTI nº 4.979, de 13 de julho de 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ia_estrategia_portaria_mcti_4-979_2021_anexo1.pdf. Acesso em: 24 set. 2021.

BRASKEN. Como o uso de redes sociais impacta nossa saúde mental. Disponível em: https://bluevisionbraskem.com/desenvolvimento-humano/como-o-uso-de-redes-sociais-impacta-nossa-saude-mental/. Acesso em: 7 jun. 2020.

BRUNO, Fernanda. *Máquinas de ver, modos de ser*: vigilância, tecnologia e subjetividade. Porto Alegre: Sulina, 2013.

CANALTECH. Secretária Virtual da Amazon fala Português. Disponível em: https://youtu.be/-l1rmRrompQ. Acesso em: 1º ago. 2020.

CECI, Mariana. Aula de longe, mas ao pé do ouvido. Disponível em: https://piaui.folha.uol.com.br/aula-de-longe-mas-ao-pe-do-ouvido/. Acesso em: 7 jun. 2020.

CONSELHO NACIONAL DE JUSTIÇA. Resolução nº 332. Disponível em: https://www.anoreg.org.br/site/wp-content/uploads/2020/08/Resolução-332-CNJ.pdf. Acesso em: 2 set. 2020.

DOBELLI, Rolf. *Die Kunst des digitalen Lebens*. München: Piper Verlag, 2019, s. 131-134.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. 2. ed. São Paulo: Revista dos Tribunais, 2020.

GAVRAS, Douglas; LINDNER, Julia. Invisíveis, 5,5 milhões de informais correm risco de perder ajuda de R$ 600. Disponível em: https://economia.uol.com.br/noticias/estadao-conteudo/2020/04/20/invisiveis-55-milhoes-de-informais-correm-risco-de-perder-ajuda-de-r-600.htm. Acesso em: 5 jun. 2020.

GOLEMAN, Daniel. *Foco*: a atenção e seu papel fundamental para o sucesso. Trad. Cassia Zanon. Rio de Janeiro: Objetiva, 2014.

HAN, Byung-Chul. O coronavírus de hoje e o mundo de amanhã. Disponível em: https://brasil.elpais.com/ideas/2020-03-22/o-coronavirus-de-hoje-e-o-mundo-de-amanha-segundo-o-filosofo-byung-chul-han.html. Acesso em: 2 jul. 2020.

HAWKING, S. W. *Uma breve história do tempo*. Trad. Cassio de Arantes Leite. Rio de janeiro: Intrínseca, 2015.

HENNING, Klaus. *Smart und digital*: wie künstliche Intelligenz unser Leben verändert. Berlin: Springer, 2019.

HERCULANO-HOUZEL, S. Herculano-Houzel, what modern mammals teach about the cellular composition of early brains and mechanisms of brain evolution. In: *Evolutionary Neuroscience*, 10.1016/B978-0-12-820584-6.00014-3, p. 349-375, 2020.

HOFFMANN-RIEM, Wolfang. Inteligência artificial como oportunidade de regulação jurídica. *Revista de Direito Público*, n. 90, v. 16, p. 11-38, nov./dez. 2019.

IDC. IDC Futurescape: Worldwide Mining 2021 Predictions. Disponível em: https://www.reportlinker.com/p03374506/IDC-FutureScape-Worldwide-Mining-Predictions.html. Acesso em: 24 set. 2021.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Internet chega a 79,1% dos domicílios do País. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/27515-pnad-continua-tic-2018-internet-chega-a-79-1-dos-domicilios-do-pais. Acesso em: 20 ago. 2020.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA). Internet no Brasil reproduz desigualdades do mundo real. Disponível em: https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=34796. Acesso em: 22 ago. 2020.

_____. Democracia digital: mapeamento de experiências em dados abertos, governo digital e ouvidorias públicas. Disponível em: http://repositorio.ipea.gov.br/bitstream/11058/10440/1/td_2624.pdf. Acesso em: 24 set. 2021.

INTERNET WORLD STATS. Internet Usage Statistics. Disponível em: https://www.internetworldstats.com/stats.htm. Acesso em: 20 ago. 2020.

LEITE, Flávia Piva Almeida. O exercício da liberdade de expressão nas redes sociais e o Marco Civil da Internet. *Revista de Direito Brasileiro*, v. 13, n. 06, 2016.

LIMA, K. R. et al. Trabalhando conceitos da neurociência na escola: saúde do cérebro e plasticidade cerebral. *Rev. Ciênc. Ext.*, v. 13, n. 2, p. 71-82, 2017.

LIMBERGER, Têmis. *Cibertransparência* – Informação pública em rede – A virtualidade e suas repercussões na realidade. 1. ed. Porto Alegre: Livraria do Advogado, v. 01, 2016.

LIN, Junquan et. al. RNA interference in glial cells for nerve injury treatment. Disponível em: https://journals.sagepub.com/doi/pdf/10.1177/2041731420939224. Acesso em: 2 ago. 2020.

MARHOUNOVÁ, Lucie. Artificial selection on brain size leads to matching changes in overall number of neurons. Disponível em: https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.13805. Acesso em: 10 ago. 2020.

MARSDEN, C. T. *Net neutrality*: towards a co-regulatory solutions. Londres: Bloomsburry Academic, 2010.

MARTINS, Vera. *O emocional inteligente*: como usar a razão para equilibrar a emoção. Rio de Janeiro: Alta Books, 2015.

MASSIS, Diana. Somos cada vez menos felizes e produtivos porque estamos viciados na tecnologia. Disponível em: https://www.bbc.com/portuguese/geral-51409523?ocid=wsportuguese.chat-apps.in-app-msg.whatsapp.trial.link1_.auin. Acesso em: 13 jul. 2020.

MENDES, Laura Schertel; MATTIUZZO, M. Discriminação algorítmica: conceito, fundamento legal e tipologia. *Revista Direito Público*, n. 90, v. 16, nov./dez. 2019.

NAÇÕES UNIDAS. Agenda 2030. Disponível em: https://nacoesunidas.org/pos2015/agenda2030/. Acesso em: 29 ago. 2020.

NISSENBAUM, Helen. *Privacy in context*: technology, policy and the integrity of social life. Stanford: Stanford University Press, 2010.

RIBEIRO, Djamila. *Pequeno manual antirracista*. São Paulo: Companhia das Letras, 2019.

RODOTÀ, Stefano. *A vida na sociedade da vigilância* – A privacidade hoje. Trad. Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

ROTHBLATT, Martine. *Virtualmente humanos*: as promessas e os perigos da imortalidade digital. Trad. Jeferson Luiz Camargo. São Paulo: Cultrix, 2016.

SARLET, Gabrielle Bezerra Sales; CALDEIRA, Cristina. O consentimento informado e a proteção de dados pessoais de saúde na internet: uma análise das experiências legislativas de Portugal e do Brasil para a proteção integral da pessoa humana. In: *Civilistica.com*, v. 8, n. 1, p. 1-27, 29 abr. 2019. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/411. Acesso em: 2 jul. 2020.

_____; MOLINARO, Carlos Alberto. Questões tecnológicas, éticas e normativas da proteção de dados pessoais na área da saúde em um contexto de *big data*. *Direitos Fundamentais & Justiça*, Belo Horizonte, a. 13, n. 41, p. 183-212, jul./dez. 2019.

SCHMIDT, Eric; COHEN, Jared. *The new digital age*: reshaping the future of people, nations and business. London: John Murray, 2014.

SCOLA, Daniel. A farra contra aposentados que a nova lei precisa corrigir. Disponível em: https://gauchazh.clicrbs.com.br/colunistas/daniel-scola/

noticia/2020/08/a-farra-contra-aposentados-que-a-nova-lei-precisa-corrigir-ckeids4u00006014y5d8w3ai9.html. Acesso em: 31 ago. 2020.

SELF, J. C.; CODDINGTON, J. A.; FOLI, K. J.; BRASWELL, M. L. Assent in pediatric patients. *Nurs Forum*, v. 52, n. 4, p. 366-376, Oct. 2017. Disponível em: https://pubmed.ncbi.nlm.nih.gov/28419461/. Acesso em: 20 jul. 2020.

SERRES, Michel. *Narrativas do Humanismo.* Trad. Caio Meira. Rio de Janeiro: Bertrand Brasil, 2015.

SILVA, Rafael Rodrigues da. MIT enumera quais foram as dez piores tecnologias inventadas no século XXI. Disponível em: https://canaltech.com.br/curiosidades/mit-enumera-quais-foram-as-dez-piores-tecnologias-inventadas-no-seculo-xxi-135573/. Acesso em: 21 ago. 2020.

_____. Brasil é o segundo país do mundo a passar mais tempo na internet. Disponível em: https://canaltech.com.br/internet/brasil-e-o-segundo-pais-do-mundo-a-passar-mais-tempo-na-internet-131925/. Acesso em: 4 jun. 2020.

SOPRANA, Paula. Diretor do Fórum Econômico Mundial defende remuneração por dado pessoal. Disponível em: https://www1.folha.uol.com.br/mercado/2019/11/diretor-do-forum-economico-mundial-defende-remuneracao-por-dado-pessoal.shtml. Acesso em: 21 ago. 2020.

TAVARES, F.; DE ARAÚJO, F.; CARVALHO, G.; VILAÇA E SILVA, K. Guia de agulhas para transdutores lineares e convexos para punção da tireoide guiada por ecografia: "ecothyrohealth". *Revista Brasileira de Ciências do Envelhecimento Humano*, 16(1), 168-170. Disponível em: https://doi.org/10.5335/rbceh.v16i1.9911. Acesso em: 29 ago. 2020.

TODOROV, Tzvetan. *Os inimigos íntimos da democracia*. Trad. Joana Angelica d'Avila Melo. São Paulo: Companhia das Letras, 2012.

TOKARNIA, Mariana. Um em cada 4 brasileiros não tem acesso à internet, mostra pesquisa. Disponível em: https://agenciabrasil.ebc.com.br/economia/noticia/2020-04/um-em-cada-quatro-brasileiros-nao-tem-acesso-internet. Acesso em: 21 ago. 2020.

UNIÃO EUROPEIA. Comission white paper on artificial intelligence. Disponível em: https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_pt.pdf. Acesso em: 2 set. 2020.

_____. Proposal for a Regulation for the European Parliament and the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts. Disponível em: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206. Acesso em: 24 set. 2021.

VELASCO, Clara et al. Quase metade dos estados não divulga raça de mortos pela polícia; dados disponíveis revelam que 3/4 deles são negros. Disponível em: https://g1.globo.com/monitor-da-violencia/noticia/2020/09/03/quase-metade-dos-estados-nao-divulga-raca-de-mortos-pela-policia-dados-disponiveis-revelam-

que-34-deles-sao-negros.ghtml?utm_source=push&utm_medium=app&utm_campaign=pushg1. Acesso em: 3 ago. 2020.

VESCE, Gabriela. Exclusão digital. Disponível em: https://www.google.de/amp/s/www.infoescola.com/sociologia/exclusao-digital/amp/. Acesso em: 8 jun. 2020.

VOLPATO, Bruno. Ranking das redes sociais 2020. Disponível em: https://resultadosdigitais.com.br/blog/redes-sociais-mais-usadas-no-brasil/. Acesso em:
6 jun. 2020.

WEIDENFELD, Nathalie; NIDA-RÜMELIN, Julian. *Digitaler Humanismus*: eine Ethik für das Zeitalter der künstlichen Intelligenz. München: Piper, 2018, s. 53.

Sobre a autora:

Gabrielle Bezerra Sales Sarlet | *E-mail:* gabriellebezerrasales@gmail.com

Advogada. Pós-Doutora em Direito pela Universidade de Hamburgo – Alemanha. Pós-Doutora em Direito pela PUC-RS. Doutora em Direito pela Universidade de Augsburg – Alemanha. Graduada e Mestre em Direito pela Universidade Federal do Ceará. Ex-Bolsista do MPI – Max Planck Institute Hamburg – Alemanha. Professora do Curso de Graduação e de Pós-Graduação *Stricto* e *Lato Sensu* em Direito na Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Especialista em Neurociências e comportamento na PUC-RS e, atualmente, encontra-se em formação no Círculo Psicanalítico Freudiano do Rio Grande do Sul. Currículo: http://lattes.cnpq.br/9638814642817946.

Data de submissão: 30 de setembro de 2021.
Data de aceite: 12 de novembro de 2021.

O Uso da Inteligência Artificial na Repercussão Geral: Desafios Teóricos e Éticos

Artificial Intelligence Uses in Repercussão Geral: Ethical and Theoretical Issues

FAUSTO SANTOS DE MORAIS[1]
Faculdade Meridional (Imed/RS).

RESUMO: O Supremo Tribunal Federal desenvolveu com inteligência artificial o programa Victor para identificar recursos e classificá-los a temas de repercussão geral já definidos pelo tribunal. O programa computacional se destaca pela sua eficiência, contribuindo para a celeridade processual e a consistência decisória de acordo com a premissa normativa definida, leia-se, a tese da repercussão geral. De outro lado, questões teóricas e éticas emergem também do seu uso. Teoricamente, a aplicação da repercussão geral pode apresentar o mesmo problema que o precedente, atraindo o risco da hipernormatização artificial com a sobrepadronização fática ou normativa. No campo da ética, surge a questão epistêmica sobre a automatização na aplicação da repercussão geral. Isso exige ações práticas e normativas para contornar os problemas aqui nominados como de raciocínio jurídico e da responsabilidade do agente.

PALAVRAS-CHAVE: Inteligência artificial; repercussão geral; hipernormatização; questão ético-epistêmica.

ABSTRACT: The brazilian Supreme Court developed with artificial intelligence the Victor program to identify and classify appellations according to topics of repercussão geral already defined by the court. The computer program stands out for its efficiency, contributing to procedural speed and decision consistency in accordance with the normative premise defined, that is, the general repercussion thesis. On the other hand, theoretical and ethical issues also emerge from its use. Theoretically, the application of general repercussion may present a problem as precedent can, attracting the risk of artificial hypernormatization with factual or normative overstandardization. In the field of ethics, an epistemic question arises about automation in general repercussion uses. This question requires practical and normative actions to deal with problems as legal reasoning and agent responsibility.

1 Orcid: http://orcid.org/0000-0002-4648-2418.

KEYWORDS: Artificial intelligence; repercussão geral; hypernormatization; ethical-epistemic matter.

SUMÁRIO: Introdução; 1 A jurisprudência brasileira como terreno fértil à utilização da inteligência artificial; 2 O uso do programa Victor e a consistência jurisprudencial brasileira; 3 A hipernormatização artificial; 4 O elemento ético na questão epistêmica pelo uso da IA no Direito; Considerações finais; Referências.

INTRODUÇÃO

A inteligência artificial (IA)[2] vem sendo empregada em longa escala no desempenho das tarefas humanas. Isso faz com que diversas áreas do saber acabem incorporando elementos da inteligência artificial para auxiliar, melhorar ou substituir, com maior eficiência, as atividades humanas. Exemplo disso é o emprego de sistemas inteligentes para o diagnóstico de doenças de forma mais precisa que o médico humano.

Daí é possível estabelecer diferentes relações entre a inteligência artificial e o Direito, delimitando campos de pesquisa próprios. Um primeiro campo de pesquisa pode ser concebido como o Direito na inteligência artificial, no qual o uso da inteligência artificial nas mais diversas tarefas humanas exige uma reflexão sobre as suas consequências jurídicas, seja em virtude de uma necessidade de regulamentação, seja pela determinação da forma pela qual o Direito pode fornecer uma resposta jurídica aos problemas, como a responsabilidade civil e penal de atos praticados por sistemas inteligentes. Um bom exemplo desse tipo de provocação à reflexão jurídica é a discussão sobre a comercialização de robôs inteligentes com a fisionomia de crianças para fins sexuais. Atualmente, tramita no congresso norte-americano um projeto de lei, o *Curbing Realistic Exploitative Electronic Pedophilic Robots* (*CREEPER), Act*, que visa proibir a importação e comercialização desses robôs.

De outro lado, inaugura-se a discussão também sobre o uso da inteligência artificial no Direito. Isso implica em pesquisar, desenvolver e refletir sobre a influência que as práticas jurídicas estão sofrendo pelo emprego da inteligência artificial. Existe um certo consenso que o emprego dos sistemas inteligentes será direcionado para auxiliar na prática de atos, como a pes-

2 Se emprega o termo inteligência artificial em um sentido amplo, compreendendo programas computacionais capazes de reproduzir alguma das habilidades humanas. Frankish e Ramsey conceituam no seu glossário Inteligência Artificial como "a tentativa de fazer computadores fazerem atividades humanas das mais variadas" (Frankish; Ramsey, 2014). Nesse sentido, atividades automatizadas e sistemas especialistas também estariam incorporados por esse conceito.

quisa, a organização, a classificação e a recuperação de dados, sejam elas particulares ou públicas. No Brasil, já existem programas responsáveis pela classificação de ações judiciais a determinados temas jurídicos repetitivos, como são os casos do Victor no Supremo Tribunal Federal e do Radar no Tribunal de Justiça de Minas Gerais. Denomina-se essa influência como a inteligência artificial no Poder Judiciário.

Para fins do presente trabalho, parte-se do emprego da inteligência artificial pelo Poder Judiciário como um fenômeno que repercute sobre a concepção de Direito como um empreendimento argumentativo. Em decorrência disso, sustenta-se que a teoria do Direito precisa revisar os seus elementos estruturais em virtude da transformação que o uso da IA no Judiciário vem produzindo.

Vale esclarecer que as questões sobre a integração da IA na teoria do Direito exigem um esforço teórico sobre a função do Direito, as suas fontes, os seus atores e a sua forma de legitimação. Esse é o pano de fundo para se colocar as questões teórica da hipernormatização artificial e ética sobre a condição epistêmica pela latente automação pelo programa Victor. Para tanto, em um primeiro momento, apresentará o contexto brasileiro frutífero para o desenvolvimento de inteligências artificial para uso no Judiciário. Em uma segunda etapa, explicar-se-á como funciona o programa Victor. Após, a indagação envolverá o fenômeno da hipernormatização artificial, para, como último tópico, levantar a questão ético-epistêmica por força da automação no Supremo Tribunal Federal.

1 A JURISPRUDÊNCIA BRASILEIRA COMO TERRENO FÉRTIL À UTILIZAÇÃO DA INTELIGÊNCIA ARTIFICIAL

Como que o Direito é aplicado? Se essa pergunta fosse realizada a um advogado ou juiz, uma possível resposta seria que o jurista deve conhecer as normas vigentes de uma determinada ordem jurídica, sendo essas as premissas a serem verificadas no caso concreto. O raciocínio jurídico envolveria, portanto, conhecer as normas vigentes e relacioná-las ao caso concreto[3]. Essa noção encontra alguma correspondência à definição de Carlos Maximiliano, para quem a aplicação do Direito consiste "no enquadrar um

3 Robert Alexy escreve o seu livro *Teoria da argumentação jurídica* vislumbrando um aporte teórico para responder quais os requisitos uma decisão judicial deveria atender – mesmo que de forma aproximada – para ser considerada juridicamente legítima (Alexy, 2005, p. 5).

caso concreto em a norma jurídica adequada" (2011, p. 5)[4]. Não parece ser diferente também da concepção de Betti sobre aplicação do Direito como "operação intelectiva, voltada a verificar e controlar se recorrem no caso os pressupostos de fato da norma" (2007, p. 12).

Essa ideia tradicional sobre a aplicação do Direito pressupõe um vasto conhecimento sobre o material legislativo e jurisprudencial vigente, com base no qual o jurista deve identificar a premissa jurídica que será utilizada para resolver o problema concreto. Atualmente, entretanto, diante do volume do material jurídico existente, é normal que o jurista realize a sua pesquisa por meio de ferramentas eletrônicas. Só para exemplificar, a identificação e o acesso ao material legislativo podem ser facilmente realizados pela ferramenta de pesquisa eletrônica Google. De forma parecida, a jurisprudência dos Tribunais também pode ser conhecida através de ferramentas de pesquisa fornecida pelos próprios *sites* dos Tribunais. É plenamente plausível, portanto, que o jurista utilize tais ferramentas eletrônicas para identificar e conhecer o material jurídico.

Além do conhecimento sobre o material jurídico, o jurista precisa produzir documentos para formalizar a aplicação do Direito. Isso pode ser feito por meio de uma petição, um parecer, uma sentença ou um acórdão. Em uma idealização, novos casos relacionados à temática deveriam ter como referência aquele documento em que a aplicação do Direito foi formalizada, a menos que o jurista seja organizado o bastante para manter um arquivo sobre as temáticas pesquisadas, o que não é impossível. De qualquer sorte, seria uma questão de eficiência e celeridade recorrer às pesquisas e aos resultados já ocorridos sobre aquela temática. Por que o Poder judiciário não pode agir da mesma maneira?

A ordem jurídica brasileira produz um vasto material legislativo e jurisprudencial. Às vezes é difícil, inclusive, estimar a quantidade de leis vigentes (Dimoulis, 2011, p. 192). É claro que esse número considera as leis produzidas em diferentes níveis federativos e normativos (Emenda Constitucional, Leis Ordinárias, Decretos, Instruções Normativas, etc.). Por sua vez, a produção jurisprudencial também possui um volume imponente. Em uma pesquisa rápida, o Superior Tribunal de Justiça tem o registro de 641

4 Poderia ser objetado que a noção apresentada não mais corresponde a forma de aplicação do Direito por força da virada linguística, em virtude da qual a unidade da *applicattio* não permitiria ao intérprete se esconder atrás de um método de aplicação, no caso, o dedutivo. Embora não se negue a problematização dos métodos na aplicação do Direito, não se pode negar a existência de um sentido usual do texto jurídico.

Súmulas e 1.048 Temas de Recursos Repetitivos; enquanto o Supremo Tribunal Federal possui 736 Súmulas, 56 Súmulas Vinculantes e 1.082 Temas de Repercussão Geral (STF, 2020a, *on-line*). Essa estrutura de jurisprudência sumulada encontra espaço também em tribunais de jurisdição comum e especial, regionais e estaduais. Conclusivamente, a quantidade de texto que serve de material à construção da premissa para o raciocínio jurídico é estravagante.

Colocada essa questão, já é possível apresentar a questão da vinculação jurisprudencial no Brasil. Embora a ordem jurídica brasileira esteja ligada pela sua tradição ao modelo da *civil law*, não se pode negar que a jurisprudência se tornou a fonte primordial do Direito[5]. O argumento é simples: advogados e juízes, quando se deparam com um caso que até então não lhe era conhecido, iniciam a sua investigação pela jurisprudência. Isso, por si só, não seria um problema.

O problema é que falta racionalidade à aplicação do Direito na ordem jurídica brasileira quando a jurisprudência dos tribunais de cúpula não é observada pelo tribunal inferior ou pelo juiz inicial. O centro da ideia é que o Judiciário está atrelado à formalidade da noção de justiça procedimental[6], conduzida pelo imperativo da igualdade. Ou seja, os casos iguais devem ser tratados de maneia igual e os casos diferentes, tratados de forma diferente, na medida da sua diferença. Primar por esse tratamento igualitário entre diferentes casos seria uma questão de consistência (Dworkin, 1977, p. 122)[7]. Aliás, característica essa primordial para uma ideia de racionalidade tanto que veio a ser uma exigência posta pelo Código de Processo Civil brasileiro de 2015[8].

5 Evidência disso é a previsão feita pelo Código de Processo Civil no art. 927, na qual devem os juízes e tribunais observarem as decisões do Supremo Tribunal Federal em controle concentrado de constitucionalidade, os enunciados de súmula vinculante, os acórdãos em demandas repetitivas, as súmulas do Supremo Tribunal Federal e Superior Tribunal de Justiça e as orientações do órgão plenário a que estiverem vinculados (Brasil, 2015, *on-line*).

6 Assume-se, nesse ponto, que o procedimento estabelecido para a aplicação do direito confere a legitimação na sua aplicação (Luhmann, 1980, p. 45), atendendo ao imperativo moral da igualdade associado à ideia de justiça.

7 Em que pese a teoria do Direito utilizar a expressão "consistência", o Código de Processo Civil brasileiro adotou a expressão "coerência" ao enunciar o dever das decisões judiciais respeitarem decisões anteriores, especialmente aquelas estabelecidas em uma jurisprudência estabilizada por enunciados sumulares.

8 O Código de Processo Civil brasileiro, em diversos momentos, faz referência às exigências de estabilidade, integridade e coerência (aqui traduzida por consistência), como se pode ver, por exemplo, no art. 926 (Brasil, 2015, *on-line*).

Contra esse requerimento de consistência, seria possível argumentar que o juiz é livre para decidir o caso de acordo com as circunstâncias concretas, promovendo a concretude do Direito, inclusive para deixar de reproduzir o entendimento fixado na cúpula do Judiciário. Também se poderia argumentar que, mesmo os recursos repetitivos, as repercussões gerais e as súmulas vinculantes terem o *status* jurídico de vinculação[9], ainda assim ao julgador haveria a possibilidade de se desviar do posicionamento consolidado.

Diante disso, parece estar cada vez mais presente a concepção decisionista sobre a aplicação do Direito, o que remonta ao modelo kelseniano de norma. Para Kelsen, a norma jurídica é um produto da vontade do julgador (2003, p. 101). Somente o julgador do caso pode expressar a sua vontade para, por meio da imputação, atribuir determinada consequência jurídica a um fato (Kelsen, 2003, p. 101). Interessa registrar aqui que a lógica jurídica no molde decisionista demanda uma imputação, em vez de uma derivação deducionista[10]. Ou seja, a decisão jurídica decorre de um ato de vontade do juiz, cabendo-lhe, inclusive, criar uma norma ou, até mesmo, deixar de aplicar a existente (Kelsen, 2011, p. 393).

Em uma palavra: a aplicação do Direito dependeria do juiz que, em uma atividade cognitiva e volitiva, identifica a premissa jurídica, vincula ela ao caso concreto e imputa a consequência jurídica. Essa atividade pode seguir ou não os padrões anteriormente definidos de acordo com a similaridade entre as circunstâncias do caso anterior e do posterior.

Sobre este contexto, duas seriam as grandes exigências à aplicação do Direito: a primeira, a capacidade de identificar o material jurídico precedente; a segunda, manter o máximo de consistência entre aquilo que já foi decidido e aquilo que precisa ser. Para ambas as exigências, a inteligência artificial (IA) seria capaz de contribuir.

Kevin Ashley aponta para os avanços da inteligência artificial voltada ao Direito, especialmente para auxiliar o jurista no exercício das suas ativi-

9 Quando se refere ao *status* jurídico de vinculação, procura expressar que o próprio estatuto processual faz a previsão de instrumentos que provocam os juízos a reconhecerem a existência de uma decisão já proferida sobre a matéria, como é o caso do instituto da reclamação, prevista no art. 988 do CPC (Brasil, 2015, *online*).

10 Kelsen afirma que a aplicação do Direito não poderia ser algo causal, ordenado pelo campo de lógica. Mas um exercício da vontade do juiz, incontrolável do ponto de vista lógico (Kelsen, 2011, p. 393).

dades[11]. Conforme ressalta Ashley, a inteligência artificial tem atualmente a capacidade de extrair informações dos textos legais e auxiliar os humanos a resolverem as questões jurídicas (2017, p. 11). Essa colaboração seria a marca de programas computacionais caracterizados pela recuperação argumentativa (*Argument Retrieval*) e pela computação cognitiva (*Cognitive Computing*)[12]. O funcionamento de programas como esses pressupõe a participação dos juristas em uma etapa de modelação do raciocínio jurídico, seja com base na legislação (*Statutory Reasoning*) ou em precedentes (*Cased-Based Reasoning*).

Quando se fala em modelação com base na legislação (*Statutory Reasoning*), o pressuposto está na possibilidade de representação lógica das regras legislativas. Isso permitiria, inclusive, que a complexidade da legislação decorrente do volume legislativo pudesse ser facilitada pela capacidade computacional da IA. Assim, seria possível incluir no sistema uma descrição fática do caso para que o programa retornasse com a regra jurídica aplicável.

Uma determinada regra legislativa específica poderia[13] ser modelada por meio de proposições lógicas, como ilustra Ashley com o exemplo da formalização proposicional no caso da Seção 354 do *Internal Revenue Code* (2017, p. 42). Isso resolveria parte da questão de identificação do material jurídico legislativo utilizado como premissa para o raciocínio jurídico.

Havendo a modelação do raciocínio com base na legislação, o material jurídico não pode ser considerado completo sem a integração do direito legislado com o direito jurisprudencial. Aliás, no contexto brasileiro, conforme já apresentado, as proposições sumulares desempenham um papel central como fonte do Direito ou elemento principal na compreensão da

11 Ashley sustenta que os programas criados para lidarem com informações decorrentes da modelação baseadas em regras ou em casos têm a capacidade de oferecer razões, as quais serviriam para realizar tarefas, como raciocinar, explicar, argumentar e predizer (Ashley, 2017, p. 32), apresentando exemplos de programas capazes de realizar tarefas inteligentes, como o raciocínio jurídico, a explicação e a argumentação.

12 Seguindo com a ideia de Ashley, a recuperação argumentativa (*Argument Retrieval*) implicaria a pesquisa e a identificação dos conceitos legais de forma automática, capazes de contribuir com a solução de um problema jurídico. Com os avanços no processamento da linguagem natural, a recuperação argumentativa seria inclusive capaz de identificar automaticamente os argumentos presentes nos textos jurídicos. Por computação cognitiva (*Cognitive Computing*), Ashley identifica a capacidade de programas em extrair possíveis soluções de um determinado corpo textuais e ranqueá-las de acordo com a sua relevância para o problema (Ashley, 2017, p. 13).

13 Está se considerando aqui o uso de um programa computacional especialista. Ashley esclarece que a especialidade do sistema seria a capacidade de aplicar regras jurídicas a fatos específicos, mediante prévia programação (Ashley, 2017, p. 8).

premissa para uma conclusão jurídica. Portanto, para além da modelação legislativa, deve-se pensar também sobre a modelação da jurisprudência como uma necessidade decorrente da ideia de consistência jurídica para a harmonização jurisprudencial.

Para a modelação da jurisprudência se deve definir qual é o objeto que se pretende representar. A literatura especializada reporta que a modelação da jurisprudência envolveria a capacidade de representar um precedente (*Case-Based Reasoning*), tendo-se aqui em mente um determinado caso concreto decidido pelo Judiciário. Isso envolveria conhecer os elementos fáticos considerados para a conclusão jurídica, representando-os como fatores que suportam a comparação entre diferentes casos jurídicos.

Poder-se-ia concluir, de forma apressada, que a integração da modelação da legislação com o precedente seria uma representação correta da estruturação do raciocínio jurídico brasileiro. Objeta-se, entretanto, que a forma de aplicação da jurisprudência no Brasil possui a particularidade própria da criação de proposições sumulares. Ao contrário dos precedentes, nos quais os elementos fáticos são determinantes para a comparação entre casos jurídicos, as proposições sumulares[14] apresentam uma característica estrutural igual à legislação. Dessa maneira, aquela definição jurisprudencial que já foi consolidada em um enunciado sumular deve ser representada nos moldes de um raciocínio estatutário (*Statutory Reasoning*).

Talvez mais importante que a forma de modelação do raciocínio jurídico nos moldes da legislação (*Statutory Reasoning*) e do precedente (*Case-Based Reasoning*) seja o emprego de técnicas de *Machine Learning*, supervisionada ou não, para o processamento de linguagem natural. Essa técnica ajudaria a identificar os padrões jurídicos na legislação, na jurisprudência e nos precedentes, sem requerer o mesmo esforço sobre a modelação do raciocínio jurídico.

A próxima seção apresentará dois programas que possuem a potencialidade de aplicação da inteligência artificial para otimizar a consistência jurisprudencial na ordem jurídica brasileira.

14 E aqui se está referindo à diversidade de enunciados de súmulas, súmulas vinculantes, teses em casos repetitivos, entre outros.

2 O USO DO PROGRAMA VICTOR E A CONSISTÊNCIA JURISPRUDENCIAL BRASILEIRA

Pelo menos dois programas de inteligência artificial já funcionam no Judiciário brasileiro, oferecendo, para além de uma maior celeridade, contributos para uma maior consistência jurisprudencial. Como dito antes, se um determinado caso já teve a sua questão de mérito jurídico definida pelos tribunais superiores e não havendo motivos pelos quais o juízo das instâncias inferiores reconheçam uma distinção (*distinguishing*), deve-se, por dever de consistência, manter o padrão decisório.

Como ficará adiante entendido, os programas apresentados não são utilizados no processo decisório. Não são os programas que definem qual a norma que vai ser aplicada ao caso. Para tanto, eles dependem de um ato decisório do tribunal para definir a norma que resolve a demanda repetitiva. Feito isso, os programas compartilham a característica de classificarem as demandas judiciais aos temas já definidos pelo tribunal. Esse procedimento que permite a identificação e classificação dos temas de demanda não possui o *status* jurídico de vinculação normativa. Todavia, do ponto de vista fático, consegue fazer uma vinculação tecnológica com o efeito normativo.

O primeiro programa a ser indicado é o Victor, no Supremo Tribunal Federal. Vale, preliminarmente, esclarecer o contexto legal processual de funcionamento do programa. De uma forma geral, cabe ao Supremo Tribunal Federal definir os temas de repercussão geral ou não, vinculando os recursos apresentados ao tribunal a esses temas. Assim, tem-se três resultados: sem repercussão, com repercussão ou, ainda, não definido. Quando o tema não possui repercussão, os recursos apresentados sobre aquela temática devem ter a sua admissibilidade negada. Se o tema possui repercussão geral, surgem duas possibilidades: já julgado, os recursos apresentados devem reproduzir o mesmo entendimento; ainda não julgado, reúnem-se os recursos sobre a mesma matéria em um grupo, aguardando o julgamento pelo caso representativo. No caso de a temática ainda não ter sido definida como de repercussão geral, os recursos sobre o mesmo assunto são reunidos em grupos e ficam aguardando essa definição.

Até hoje foram definidos 1.082 temas de repercussão geral. Desses, em 338 casos foi definido não haver repercussão geral. Nos demais 730 casos em que a repercussão geral foi reconhecida, 419 já tiveram a sua temática julgada definitivamente, havendo 311 temas pendentes de julgamento (STF, 2021a, *on-line*).

Para lidar com essa realidade, de um grande de volume de recursos sobre o mesmo tema e com a possibilidade de reuni-los em grupo para um julgamento temático, idealizou-se o desenvolvimento de um programa de IA, o que foi realizado por meio do projeto Victor. Esse projeto teve início no ano de 2018 através de uma parceria entre a Universidade de Brasília (UnB) e o Supremo Tribunal Federal. O projeto tinha como objetivo reconhecer os padrões nos processos jurídicos relativos aos julgamentos de repercussão geral no Supremo Tribunal Federal (Peixoto, 2020, p. 5).

Importante esclarecer que os recursos encaminhados ao Supremo Tribunal Federal formavam um arquivo de documentos digitalizados ou produzidos em formato Word convertido em PDF. Com base nesses documentos, os recursos deveriam ser classificados a um determinado tema de repercussão geral. Essa era a prática humana da secretaria judiciária do tribunal. Estima-se que essas funções de identificação do tipo de tema por meio dos documentos pelos servidores humanos do tribunal levavam uma média de tempo entre 30 a 40 minutos. Descobriu-se que a acurácia nessa prática era de 75%. Isso quer dizer que 25% do trabalho realizado deveria ser realizado novamente.

Com o emprego de técnicas de *machine learning*, a pesquisa em julho de 2019 já constituía uma base de dados (*dataset*) de 200 mil processos e com a checagem de 14 mil processos por meio do programa computacional. Para além do volume de processos envolvidos, chama a atenção os resultados obtidos: quanto ao tempo, o programa conseguia realizar a mesma atividade de um servidor humano em um lapso temporal entre 4 a 5 segundos; quanto à acurácia, o programa apresentou um êxito de 91% de correção no desempenho das atividades.

Vale registrar que nem todos os 1.068 temas de repercussão geral foram inicialmente objeto de escrutínio pelo programa Victor. Ao longo do tempo, temas estratégicos para a corte foram escolhidos em virtude do volume processual que envolviam. No ano de 2017, os temas escolhidos se relacionavam com a definição dos juros moratórios aplicados às condenações contra o Estado Brasileiro (Tema 810), a diferença salarial específica de 47,11% devida aos servidores federais (Tema 951) e a possibilidade de aplicação do tema constitucional à conversão de licença-prêmio em pecúnia (Tema 975). Para o ano de 2018, os temas escolhidos foram sobre a não incidência de verba previdenciária sobre remuneração extraordinárias dos servidores públicos (Tema 163) e a falta de observação nos tribunais inferiores do julgamento pela maioria dos seus integrantes no caso de reco-

nhecimento de inconstitucionalidade de forma difusa (Tema 739). Por fim, em 2019, os temas estratégicos envolviam questões sobre a violação do contraditório e da ampla defesa pelos tribunais de instância inferior (Tema 660), a necessidade de fundamentação das decisões judiciais (Tema 339), a impossibilidade de recurso ao Supremo Tribunal Federal dos recursos decorrentes dos Juizados Especiais Civis (Tema 800) (Peixoto, 2020, p. 10).

Para sintetizar o funcionamento do programa Victor, pode-se indicar a prática de duas ações pelo programa: a primeira, com a identificação dos documentos recursais, reunindo-os de acordo com a mesma temática; a segunda, o programa associa os recursos aos temas de repercussão geral definidos pelo Supremo Tribunal Federal. Para essa segunda ação, devem os Ministros do Supremo Tribunal Federal definirem se o tema é de repercussão geral ou não e, também, deliberar como o tema deve ser resolvido. Logo, a definição do tribunal sobre o tema jurídico envolvido segue uma lógica parecida com o funcionamento dos precedentes na *common law*. A distinção, todavia, é que, vez de promover um juízo de analogia entre os casos, o tribunal define uma proposição sumular. Essa seria a premissa normativa jurisprudencial para o funcionamento do programa Victor.

Para fins dessa reflexão, assume-se que a identificação dos recursos em temas agrupados bem como a associação desses grupos aos temas de repercussão geral definidos pelo Supremo Tribunal Federal são atos cognitivos. De outro lado, a definição do resultado deliberativo sobre os temas de repercussão geral, ou seja, qual o resultado jurídico para aquela situação, é um ato decisório-valorativo, próprio dos Ministros componentes do tribunal.

Com isso, parece ser difícil negar que o programa Victor oferece maior agilidade organizacional ao Poder Judiciário, ao mesmo tempo em que garante a otimização da consistência jurisprudencial entre os recursos interpostos junto ao STF e os temas de repercussão geral.

A iniciativa mencionada corresponde tecnologicamente às atividades do jurista na aplicação do Direito. O programa consegue, primeiro, identificar e processar os textos jurídicos, estabelecendo uma consistência entre as informações existentes. No segundo estágio, o programa é capaz de identificar se o caso concreto deveria ser categorizado como uma das classes associadas aos temas de repercussão geral escolhidos.

Aceitar essa ideia seria admitir a automatização do Direito brasileiro e a dispensabilidade do jurista ser humano? Embora a dispensabilidade

não seja provável, especialmente diante do antropocentrismo jurídico[15], entende-se que as regras do Código de Processo Civil permitem a adoção de programa de IA capazes de associar casos a temas predefinidos, algo que satisfaz a noção guia de prazo razoável de tramitação processual.

Registra-se, ainda, que, do ponto de vista teórico, os programas que apresentam as características computacionais de programação cognitiva e recuperação argumentativa[16] seriam capazes de auxiliar o jurista na identificação do material jurídico pertinente e aperfeiçoar a consistência entre as decisões judiciais.

3 A HIPERNORMATIZAÇÃO ARTIFICIAL

A *hipernormatização artificial* é o fenômeno que ocorre quando os elementos fáticos ou normativos considerados em uma decisão jurídica não são adequadamente formalizados em uma premissa para o uso por um programa de inteligência artificial. Por conta disso, a IA passa a operar com uma *premissa jurídica artificial* que é progressivamente hiperestabilizada diante do funcionamento eficaz do programa.

A adjetivação de artificialidade na premissa é pejorativa, apontando para o problema de haver um significado mais amplo que foi artificialmente reduzido para permitir a sua operacionalização. É claro que o ato de formalização da uma decisão exige uma redução de significado e apreensão do seu contexto significativo. Todavia, sustenta-se que no caso da hipernomatização artificial essa redução prejudica a representação computacional do julgado.

A questão da *hipernormatização artificial* ganha relevância para a prática argumentativa jurídica no Brasil, visto que, primeiro, vive-se uma fase de implementação de programas de IA responsáveis pela identificação, classificação e vinculação de demandas judiciais aos temas de repercussão geral e de causas repetitivas. Segundo, os tribunais superiores extraem teses dos julgamentos, como se fossem artigos de lei, tomados atomicamente como precedentes a serem obedecidos. Nesse ponto, a tese estabelecida se

15 Argumento pelo qual a prestação jurisdicional somente pode ser prestada pelo ser humano, que, em consideração aos objetivos do trabalho, não será aqui discutida.

16 Kevin Ashley apresenta, na sua *Artificial Intelligence And Legal Analytics: New Tools for Law Practice in the Digital Age*, diversas aplicações de programas de inteligência artificial no Direito com as finalidades de recuperação da informação jurídica (*conceptual legal information retrieval*) e de computação cognitiva (*cognitive computing*) (Ashley, 2017, p. 11).

descola do seu contexto de aplicação, ganhando uma autonomia significativa e informacional.

De uma forma geral, a decisão jurídica que define as teses aos temas indicados vem servindo de premissa operacional que conduz o funcionamento da IA. Isso pode ser visto tanto na experiência do Victor no Supremo Tribunal Federal e do Radar pelo Tribunal de Justiça de Minas Gerais.

Se, por um lado, é evidente que essas IAs são eficientes; de outro, fica a dúvida quanto ao espaço para manifestação do fenômeno da hipernormatização artificial. Isso pode acontecer em decorrência de fatores normativos ou fáticos, como se verá a seguir.

Os *fatores normativos* ocorrem quando se determina uma regra que sofreu modificações interpretativas pelo tribunal ou, até mesmo, o entendimento firmado não encontra mais consenso nas decisões atuais. Portanto, a regra continua sendo utilizada na sua formulação originária, embora seja mais essa a prática judiciária. Dois exemplos podem ajudar a esclarecer.

Os elementos normativos podem ocorrer quando o tribunal, ao longo do tempo, modificou a sua interpretação sobre determinado posicionamento registrado em um preceito sumular (súmula ou tema repercussão geral) ou, até mesmo, esse entendimento deixou de ser um consenso. Exemplos disso são as Súmulas nº 343 e nº 347 do STF (STF, 2021b, *on-line*).

A Súmula nº 343 do STF, aprovada em 13 de dezembro de 1963, tem como seu enunciado: "Não cabe ação rescisória por ofensa a literal disposição de lei, quando a decisão rescindenda se tiver baseado em texto legal de interpretação controvertida nos tribunais". Apesar disso, com o advento da Constituição Federal de 1988, boa parte da doutrina e da jurisprudência passou a admitir a propositura de ação rescisória contra decisões transitadas em julgados no caso de violação literal à Constituição (Cruz e Tucci, 2014, *on-line*). Portanto, o enunciado literal da súmula não poderia simplesmente ser considerado por uma IA sem a devida limitação do seu alcance significativo às violações ao texto constitucional.

A Súmula nº 347, aprovada em 13 de dezembro de 1963, também apresenta um problema de hipernormatização artificial por elemento normativo. O enunciado aprovado determina que "o Tribunal de Contas, no exercício de suas atribuições, pode apreciar a constitucionalidade das leis e os atos do Poder Público". Inobstante o texto parecer não gerar dúvidas, a atual jurisprudência no Supremo Tribunal Federal registra decisões contra-

ditórias sobre tal possibilidade, havendo uma falta de consenso na própria corte em relação à temática.

Existem ainda os elementos fáticos. Nesse caso, o enunciado formulado pelo tribunal deixa de registrar um elemento fático determinante, o que deturpa o sentido da norma no momento da aplicação em virtude de tal ausência. A Súmula Vinculante nº 11 e a fixação da tese no RE 494.601/RS são exemplos esse problema.

Em 22 de agosto de 2008, o STF publicou o enunciado da Súmula Vinculante nº 11, registrando textualmente que "só é lícito o uso de algemas em casos de resistência e de fundado receio de fuga ou de perigo à integridade física própria ou alheia, por parte do preso ou de terceiros, justificada a excepcionalidade por escrito [...]" (STF, 2020b, *on-line*). Em referência ao julgamento que deu vazão à súmula, compreende-se que a expressão "justificada a excepcionalidade por escrito" dizia respeito à necessidade da autoridade justificar concretamente, diante das circunstâncias do caso concreto, o uso das algemas por conta da conduta realmente perigosa do preso. Todavia, o STF, após a publicação do enunciado, legitimou o uso das algemas no RHC 102.962/MG (STF, 2011, *on-line*) quando a justificação apresentada pela autoridade foi meramente formal – ficou registrado que o uso foi necessário "para assegurar a integridade física doas agentes de polícia e do próprio acusado". Ou seja, essa mesma justificativa poderia ser utilizada em qualquer outro caso.

A outra situação de hipernormatização fática pode ser observada na fixação da tese no RE 494.601/RS (STF, 2019, *on-line*), em que estava em jogo a proibição do sacrifício de animais em rituais de matriz africana. O julgamento foi precedido de audiências públicas, nas quais diversos especialistas ouvidos informaram duas questões fáticas consideradas pela maioria dos Ministros, quais sejam: a) os animais sacrificados não sofriam; b) a carne dos animais era ingerida no ritual. Contudo, no momento de registro da tese, esses dois elementos deixaram de contar no texto registrado, como se pode observar: "É constitucional a lei de proteção animal que, a fim de resguardar a liberdade religiosa, permite o sacrifício ritual de animais em cultos de religiões de matriz africana".

Verifica-se, portanto, que a formalização textual sumular deve procurar representar todos os elementos fáticos determinantes para a incidência da regra sob pena de haver uma sobrepadronização significativa.

Identificados os contornos do fenômeno, cabe apresentar a deficiência metodológica na compreensão da jurisprudência como fonte do Direito, com repercussão na forma de modelação computacional da ontologia jurídica[17], como provável causa para a hipernormatização. Para ser mais claro: o problema da hipernormatização artificial estaria mais centrado na incapacidade de o jurista brasileiro compreender que o caráter normativo da jurisprudência não pode ser aprisionado por um singelo enunciado textual, mas depende de uma reconstrução da decisão, nos moldes do seu contexto significado em comparação com as similitudes fáticas de um novo caso (Streck, 2013, p. 328). Essa lógica aproximaria a aplicação da jurisprudência com a metodologia dos precedentes da *common law*.

Tal condição fica clara quando se verifica que no Brasil há uma cultura própria para a aplicação da jurisprudência como fonte do Direito, especialmente pela existência de institutos jurídicos, tais quais as súmulas dos tribunais, as teses dos recursos repetitivos e da repercussão geral, bem como as súmulas vinculantes do STF. Isso porque, ao contrário do que acontece tradicionalmente na *common law* com a fixação de um caso decidido como *holding*, para que no seu contexto venha a ser considerado como paradigmático no futuro, no Brasil o tribunal acaba por criar um enunciado textual que é aplicado como se fosse uma regra jurídica. Para tanto, o enunciado sumular isolado é elevado à premissa inicial normativa para um silogismo com o caso concreto.

Essa metodologia de aplicação do Direito influencia a forma pela qual a modelação computacional da ontologia[18] jurídica deve ser realizada. Como esclarece Ashley, a função da ontologia é prover um vocabulário conceitual para a representação do conhecimento que a IA pode processar (2017, p. 173). No caso da aplicação do Direito, duas podem ser as formas de modelagem, como será visto a seguir.

A primeira é o tipo de raciocínio com base em casos (*Cased-based Legal Reasoning*). Essa técnica representa o conhecimento sobre os fatos e as similaridades jurídicas entre casos, aproximando-os aos conceitos jurídicos (Ashley, 2017, p. 77). Criada a ontologia, a IA pode comparar o problema e os fatos, indicando a similaridade entre eles e sugerindo determinada

17 A noção de ontologia é empregada no sentido da explicitação formal e especificação geral de propriedades conceituais e as relações entre as entidades de um dado domínio (Wyner, 2008, p. 370).

18 Emprega-se ontologia jurídica para indicar a forma dominante de percepção do fenômeno jurídico em consideração ao conhecimento na área do Direito, ao raciocínio e à argumentação jurídica.

consequência jurídica. A comparação entre os casos e os fatos relevantes reproduz uma metodologia de aplicação do Direito típica da *common law*.

A segundo é o tipo de raciocínio legislativo (*Statutory Reasoning*) (Ashley, 2017, p. 32). Partindo do pressuposto que as regras legislativas podem ser representadas de maneira lógica e que a IA tem a capacidade de deduzir, seria necessário apenas o *input* do suporte fático no programa para que fosse possível identificar se as condições para a aplicação da regra teriam sido preenchidas (Ashley, 2017, p. 39).

No caso brasileiro, se poderia pensar equivocadamente que a modelação deveria se conformar ao tipo de raciocínio com base em casos, por se tratar da aplicação da jurisprudência como fonte do Direito. Entretanto, tal impressão está equivocada. Da maneira pela qual os enunciados sumulares concebidos e aplicados, estar-se-ia diante de uma ontologia típica do raciocínio legislativo. Portanto, a IA precisa modelar os enunciados jurisprudenciais como se fossem premissas cujas condições poderiam ser verificadas no caso concreto mediante um raciocino dedutivo.

4 O ELEMENTO ÉTICO NA QUESTÃO EPISTÊMICA PELO USO DA IA NO DIREITO

Viu-se até então que o programa de IA vem sendo utilizado pelo STF com objetivo principal de conferir maior celeridade e eficiência à resolução dos recursos extraordinários correlacionados a temas de repercussão geral. Não se deve ignorar também a capacidade do programa também contribuir com consistência na aplicação do Direito.

Em um primeiro momento, exaltou-se a preocupação que se deve ter sobre uma possível hipernormatização artificial e como a modelação do raciocínio jurídico deve se atentar diante de sobrepadronizações fáticas e normativas que podem desvirtuar a norma construída para a resolução de uma demanda com repercussão geral.

Resta ainda, e este é o objetivo da presente seção, revolver a prática jurídica na aplicação da repercussão geral pelo programa Victor para tornar visível o elemento ético pelos potenciais riscos na utilização da tecnologia para identificar e classificar os recursos de maneira automatizada. Uma razão para temer essa automatização é a tradição da prática judiciária de julgamento por lotes de recursos, o que, na prática, já faz a aplicação de uma decisão para diversos processos, em evidente atividade automatizada. Como seria possível garantir que essa prática não seja estendida ao funcionamento do programa Victor?

Uma tal automatização implica eticamente em uma questão epistêmica diante de problemas sobre a forma do raciocínio jurídico e o seu respectivo agente. Mesmo que para fins analíticos esses problemas possam ser apresentados separados, na prática eles estão integrados.

O problema da forma de raciocínio está em um âmbito sobre a natureza do conhecimento jurídico. Isso também poderia ser considerada como hermenêutica, eis que a compreensão sobre o papel e o funcionamento do Direito, o seu caráter institucional, não consegue ser totalmente representado por um uso analítico da linguagem. A tese aqui é que o conhecimento jurídico possui uma base linguística performativa (Searle, 2010, p. 283). Quando um recurso é negado ou provido no tribunal, não se estaria apenas conferindo mais eficiência ao procedimento judicial, mas também permitindo ou não a restrição de um direito fundamental. Nesse aspecto, o julgador assume o dever moral de estar consciente sobre as consequências da sua decisão em outro sujeito moral. Ou seja, a identificação e classificação de um recurso como de repercussão geral pelo programa computacional não são capazes de entender as condições performativas decorrentes da sua ação específica, o que pode ser a restrição de um direito fundamental.

Transitar nessa dimensão performativa exige do agente uma condição de intencionalidade (Searle, 1980, p. 423), assumindo-se aqui que a prática jurídica incorpora uma pretensão de correção (Alexy, 2020, p. 42), que capacita o jurista na avaliação quanto à correção da sua decisão diante de imperativos, tais quais as normas postas democraticamente e a proteção aos direitos fundamentais.

O programa computacional, em um processo de automatização na classificação, não age com a consciência para tematizar o significado daquilo que está fazendo (Searle, 1980, p. 420), muito menos consegue projetar as consequências das suas ações diante de outros sujeitos morais. Aquilo que o computador é capaz de fazer está circunscrito a uma operação de cálculo ausente de significação, visto que significar é um ato contextualmente dependente (Liao, 2020, p. 161). Não há, portanto, uma agente capaz de perceber eventuais riscos originários de uma visão contextual para além da operação de cálculo preestabelecida.

Isso leva à discussão sobre o problema do agente.

O problema do agente decorre do dever moral de responsabilidade do julgador. O julgador é obrigado a se engajar na prática interpretativa que o Direito é (Dworkin, 1985, p. 159), sendo seu dever moral oferecer a úni-

ca resposta correta (Dworkin, 1977, p. 282). Isso exige que cada caso seja apreciado e analisado com o mesmo engajamento, o que, diante da já exigente automatização humana, pode ser indevidamente sobrepadronizado. Quando essa prática é representada computacionalmente, tem-se, como se viu, o fenômeno da hipernormatização artificial.

A forma de resolver essa questão é tornar eficaz a Resolução nº 332/2020 do Conselho Nacional de Justiça para garantir o direito de supervisão humana para todos os afetados pelo uso da tecnologia (CNJ, 2020, *on-line*). Sendo realista, embora seja pouco provável que toda a decisão automatizada seja objeto de revisão humana, deve-se garantir institucionalmente um espaço para questionamento diante de potencial equívoco, com a previsão regimental pelo tribunal de uma instância revisória no caso de impugnação pelo interessado.

Para tanto, se fazem necessárias exigências éticas, como a explicabilidade e a transparência. Esses dois imperativos parecem ser diretrizes éticas recorrentes quando se trata sobre as exigências éticas na utilização de IA, como pode ser visto, por exemplo, na Carta Europeia de Ética sobre o uso da Inteligência Artificial em Sistemas Judiciais e seu Ambiente (CEPEJ, 2018, *on-line*). Nesse quesito, a primeira lei de sopesamento teorizada por Alexy como forma de proteção eficiente aos direitos fundamentais (2008, p. 167) deve ganhar uma roupagem própria na orientação sobre a aplicação das diretrizes éticas, podendo ser formulada da seguinte forma: quanto maior for o grau de potência lesiva aos direitos fundamentais da tecnológica utilizada, maior deve ser a extensão da clareza sobre o seu uso e a explicação sobre o seu funcionamento.

Isso não parece ser difícil de garantir no caso do programa Victor, sugerindo-se duas medidas para atender às exigências da explicabilidade e transparência. A primeira, a publicização de esclarecimentos no *website* do tribunal sobre o funcionamento do programa de forma simples para o público em geral, disponibilizando, também, estudos científicos que documentem o desenvolvimento do programa. Até o momento de finalização do presente escrito não foi possível localizar algum estudo científico. A segunda, o alerta ao jurisdicionando quanto ao uso do programa computacional na decisão do seu caso, inclusive fornecendo o respectivo caminho para acesso de maiores informações, conforme sugerido na primeira ação.

Para sintetizar o argumento sobre a questão epistêmica na ética paro o uso da IA no Judiciário, tanto o problema sobre o raciocínio quanto ao

agente pode ser contornado mediante medidas normativas que assegurem aos jurisdicionados a possibilidade de revisão de uma decisão judicial automatizada. Tais medidas importam em conferir explicabilidade e transparência sobre o funcionamento da tecnologia aplicada ao seu caso, especialmente prevendo regimentalmente no tribunal uma instância revisória de potencial decisão equivocada.

CONSIDERAÇÕES FINAIS

A integração da inteligência artificial ao Direito exige uma revisão de diversos pressupostos teóricos da teoria do Direito e da teoria da argumentação jurídica. Como se objetivou neste trabalho, foi apresentado o fenômeno da *hipernormatização artificial* como uma questão decorrente da aplicação de inteligência artificial no Direito.

A hipernormatização jurídica pode ser um problema do uso da AI no Direito, exigindo do jurista uma atenção sobre a forma pela qual a programação computacional é realizada com base em decisões jurídicas sumuladas. Essa atenção envolve argumentar sempre privilegiando os fatores jurídicos e fáticos determinantes para a formalização de premissas que serão posteriormente utilizadas para a aplicação de forma automatizada do Direito pela IA.

Há o risco de se sustentar que as causas desse efeito decorrem exclusivamente por questões tecnológicas. Entende-se, todavia, que tal posição é um equívoco. O desenvolvimento da IA no Judiciário, especialmente levando em conta os exemplos das IAs indicadas no texto, tem a característica de reproduzir a rotina dos tribunais para identificar e classificar as demandas judiciais, tomando como seu pressuposto uma determinada visão sobre o fenômeno jurídico. Essa visão é um elemento da teoria do Direito e deve ser compreendida como a causa para a hipernormatização artificial.

Por outro lado, a prevenção sobre uma hipernormatização artificial demanda ações sob as perspectivas jurídica e computacional. No campo jurídico, mostra-se necessário um aprimoramento teórico sobre a jurisprudência como fonte do Direito. Especialmente, de que maneira a construção de teses em demandas repetitivas, como premissas jurídicas ao processamento pela máquina, devem reproduzir os fatores normativos e fáticos mais relevantes dos julgamentos. Sob a perspectiva computacional, mesmo considerando a possibilidade de modelação da ontologia jurídica para orientar atos classificatórios, deve-se investir em uma compreensão transdisciplinar na

recuperação das informações jurídicas por meio de uma maior aproximação às teorias e técnicas de representação e gerenciamento do conhecimento.

Ainda sob uma perspectiva ética, a questão epistêmica sobre a forma do raciocínio jurídico e o agente responsável devem ser preocupações com consequências normativas, fazendo-se exigir medidas, como: a identificação da tecnologia para solução do caso, a explicação simplificada e também científica sobre o funcionamento do recuso ,tecnológico bem como a previsão regimental de uma instância revisora em caso de erro, mesmo quando a tecnologia for utilizada como suporte à decisão, como é o caso do programa Victor no STF.

REFERÊNCIAS

ALEXY, Robert. A Non-positivistic Concept of Constitutional Rights. *International Journal for the Semiotics of Law*, v. 33. n. 1. p. 35-46, 2020.

_____. *Teoria da argumentação jurídica*. 2. ed. São Paulo: Landy, 2005.

_____. *Teoria dos direitos fundamentais*. São Paulo: Malheiros Editores, 2008.

ASHLEY, Kevin D. *Artificial Intelligence and Legal Analytics*: new tools for law practice in the digital age. Cambridge: Cambridge University Press, 2017.

BRASIL. Lei nº 13.105. Código de Processo Civil. Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm. Acesso em: 5 maio 2020.

CEPEJ – Comissão Europeia para a Eficácia da Justiça. Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente, 2018. Disponível em: https://rm.coe.int/carta-etica-traduzida-para-portuguesrevista/168093b7e0. Acesso em: 10 jun. 2020.

CNJ – Conselho Nacional de Justiça. Resolução n 332/2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 4 nov. 2020

CRUZ E TUCCI, José Rogério. Súmula nº 343 do STF viabiliza o caminho da ação rescisória. *Conjur*. Disponível em: https://www.conjur.com.br/2014-jun-17/sumula-343-stf-viabiliza-caminho-acao-rescisoria. Acesso em: 11 jun. 2020.

DWORKIN, Ronald. *A Matter of Principle*. Cambridge: Harvard University Press, 1985.

_____. *Taking Rights Seriously*. Cambrdge: Harvard University Press, 1977.

FRANKISH, Keith; RAMSEY, William M. (Org.). *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press, 2014.

LUHMANN, Niklas. *Legitimação pelo procedimento*. Brasília: Editora Universidade de Brasília, 1980.

PEIXOTO, Fabiano Hartmann. Projeto Victor: relato do desenvolvimento da inteligência artificial na repercussão geral do Supremo Tribunal Federal. *Revista Brasileira de Inteligência Artificial e Direito*, v. 1, n. 1, p. 1-20, 2020.

SEARLE, John R. *Making The Social World*: the Structure of Human Civilization. Oxford: Oxford University Press, 2010.

_____. Minds, Brains, and Programs. *The Behavioral and Brain Sciences*, v. 3, p. 417-424, 1980.

STRECK, Lenio Luiz. *Jurisdição constitucional e decisão jurídica*. 3. ed. São Paulo: Revista dos Tribunais, 2013.

STF – Supremo Tribunal Federal. Teses de Repercussão Geral, 2021a. Disponível em: http://stf.jus.br/portal/jurisprudenciaRepercussao/abrirTemasComTesesFirmadas.asp. Acesso em: 20 set. 2021.

_____. Súmulas do STF, 2021b. Disponível em: http://www.stf.jus.br/portal/cms/verTexto.asp?servico=jurisprudenciaSumula&pagina=sumula_301_400. Acesso em: 11 jun. 2020.

_____. Relatório das atividades 2019. Brasília: [s.n.], 2020a. Disponível em: http://www.stf.jus.br/arquivo/cms/centralDoCidadaoAcessoInformacaoGestaoEstrategica/anexo/2020_01_24_13.08_RelatoriodeAtividades2019_completo.pdf.

_____. Súmulas Vinculantes, 2020b. Disponível em: http://www.stf.jus.br/portal/cms/verTexto.asp?servico=jurisprudenciaSumulaVinculante. Acesso em: 11 jun. 2020.

_____. Sacrifício de animais em rituais religiosos de matriz africana (RE 494.601/RS), DJe 19.11.2019. Disponível em: https://jurisprudencia.stf.jus.br/pages/search/sjur414970/false. Acesso em: 11 jun. 2020.

_____. Uso de algemas (RHC 102.962/MG), 2011. Disponível em: https://jurisprudencia.stf.jus.br/pages/search/sjur187260/false. Acesso em: 11 jun. 2020.

WYNER, Adam. An Ontology in OWL for legal case-based reasoning. *Artificial Intelligence and Law*, v. 16, n. 4, p. 361-387, 2008.

Sobre o autor:

Fausto Santos de Morais | *E-mail:* faustosmorais@gmail.com

Doutor e Mestre em Direito Público (Unisinos). Editor Chefe da Revista Brasileira de Direito – Qualis A1. Editor Chefe e Conselheiro Editorial da Editor Executivo da Revista Brasileira de Inteligência Artificial e Direito (RBIAD), ISSN 2675-3146. Membro Fundador da Associação Ibero-Americana de Inteligência Artificial e Direito (AID-IA). Docente da Escola de Direito e do Programa de Pós-Graduação Estrito Senso – Mestrado em Direito, da Faculdade Meridional (IMED/Passo Fundo/RS). Pesquisador com apoio da Fundação Meridional.

Data de Submissão: 30 de Setembro de 2021.
Data de Aceite: 10 de Janeiro de 2022.

"Sob Controle do Usuário": Formação dos Juízes Brasileiros para o Uso Ético da IA no Judiciário

"Under User Control": Training of Brazilian Judges for the Ethical Use of AI in the Judiciary

EUNICE M. B. PRADO[1]
Escola Nacional de Formação e Aperfeiçoamento de Magistrados (Enfam).

LUCIANE A. CORRÊA MÜNCH[2]
Escola Nacional de Formação e Aperfeiçoamento de Magistrados (Enfam).

MÁRCIA A. CORRÊA UGHINI VILLARROEL[3]
Instituto Federal do Rio Grande do Sul (IFRS).

RESUMO: Este artigo investiga a capacitação dos juízes brasileiros para exercer controle ético sobre ferramentas de inteligência artificial (IA) aplicadas à atividade jurisdicional. O Judiciário brasileiro vem investindo fortemente no uso e desenvolvimento de ferramentas de IA, contando com 64 sistemas de IA em uso ou desenvolvimento nos diversos tribunais. A Resolução nº 332/2020, que trata da ética, transparência e governança na produção e no uso de IA no Judiciário, adotou, entre outros, o princípio do controle do usuário, que estabelece direitos e deveres ao usuário quanto ao controle ético da IA. Assim, cabe ao juiz, como usuário interno, exercer tal controle ao utilizar ferramenta de IA no desempenho da atividade jurisdicional. Porém, para fazê-lo, precisa conhecer o poder-dever que lhe cabe, e estar capacitado a exercê-lo. Utilizando-se de metodologia de pesquisa quantitativa do tipo *survey*, apurou-se que a grande maioria dos magistrados brasileiros desconhece os termos da Resolução nº 332/2020; não se considera preparada para exercer controle ou supervisão de ferramentas de IA, nem tampouco se qualificou para tanto nos últimos três anos. Considerando que já há ferramentas de IA em uso pelo Judiciário, conclui-se pela necessidade de refletir sobre a política até então adotada, sugerindo que os investimentos na área sejam também direcionados à inclusão ou ampliação da capacitação dos magistrados para exercer o controle esperado.

1 Orcid: http://orcid.org/0000-0001-7221-9341.
2 Orcid: http://orcid.org/0000-0002-5984-2769.
3 Orcid: http://orcid.org/0000-0002-6676-3698.

PALAVRAS-CHAVE: Ética; inteligência artificial; magistratura; Princípios de Bangalore.

ABSTRACT: This article deals with brazilian judges' ability to perform ethical control over artificial intelligence (AI) tools applied to the judicial function. The Brazilian Judiciary has strongly invested in use and development of AI tools, with 64 AI systems currently being used or developed in various courts. Resolution nº 332/2020, which regulates ethics, transparency and governance in the production and use of AI in the Judiciary, established, among others, the principle of user control, which entails rights and duties to the user regarding the ethical control of AI. Thus, as an internal user, the judge must perform such control when using AI tools in the judicial function. However, in order to do it, they must be informed of this power-duty, and be trained to perform it. The findings of a quantitative survey research revealed that the vast majority of brazilian judges has no knowledge of Resolution nº 332/2020's contents; neither feel able to perform control or supervision of AI tools, nor has qualified for this in the past three years. Considering that the Judiciary is already using AI tools, this article points out to the need to reflect upon the current policies, suggesting that the investments in the field should also include training into the development of abilities to perform such control.

KEYWORDS: Ethics; artificial intelligence; magistracy; Bangalore Principles.

INTRODUÇÃO

À semelhança do que vem ocorrendo em diferentes esferas do setor público no Brasil e no mundo, o Poder Judiciário brasileiro passa por importante transformação digital, buscando adotar e desenvolver tecnologias que contribuam para o melhor desempenho de suas funções. Nessa jornada, já ultrapassou as fronteiras da digitalização processual para incorporar ferramentas mais sofisticadas, como sistemas de *business intelligence*, robôs, assistentes virtuais e cortes digitais.

A mais nova fronteira que vem sendo desbravada pelo Poder Judiciário é a da inteligência artificial (IA), tecnologia com potencial de impactar significativamente os julgamentos e as relações de trabalho na instituição, especialmente diante do quadro de um Judiciário congestionado e cuja força de trabalho vem sendo paulatinamente reduzida. Pesquisa realizada pela Fundação Getúlio Vargas (2020, p. 69) relata a existência de 64 projetos de IA em tribunais brasileiros, voltados ao atendimento de várias necessidades típicas da função jurisdicional, tais como: sugestões de redação de minutas, classificação de petições e processos, estimativa de probabilidade de reversão de decisões, entre outras. Diante de tamanho investimento em IA,

cabe perguntar: Estão os magistrados brasileiros preparados para os desafios trazidos por essa tecnologia?

Em 2020, o Conselho Nacional de Justiça (CNJ) baixou a Resolução nº 332/2020, que "dispõe sobre a ética, a transparência e a governança na produção e no uso de inteligência artificial no Poder Judiciário". Naquela Resolução, inspirada em grande parte pela Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente (CEPEJ, 2018), adotou-se o princípio "sob controle do usuário", com vistas a "garantir que os utilizadores sejam agentes informados e controlem as suas escolhas" (FGV, 2020, p. 20).

Ao mesmo tempo em que se trata de uma tecnologia promissora e revolucionária, a IA também traz grandes riscos que não podem ser ignorados, dado seu potencial de causar lesões a direitos. Tais riscos se revestem de especial preocupação em se tratando de sistemas de IA para uso do Poder Judiciário, um dos três Poderes do Estado Democrático de Direito instituído pela Constituição Federal, a quem cabe apreciar lesão ou ameaça a direito, nos termos dos arts. 1º, 2º e 5º, XXXV (Brasil, 1988). Conforme Peixoto (2020), o uso da IA no Judiciário traz riscos tais como: redução de direitos fundamentais, detrimento do devido processo legal, quebra da paridade de armas, violação da privacidade, redução do combate à discriminação, fragilização democrática e enfraquecimento da cidadania.

Um dos exemplos mais conhecidos e que melhor ilustra os riscos trazidos pelo uso da IA no sistema judicial é o da ferramenta norte-americana Compas, que na seara criminal criou um perfil de gerenciamento de infratores para predizer a chance de reincidência, invariavelmente concluindo que réus negros eram mais propensos a reincidir do que brancos, como analisado por Kehl e Kessler (2017). É justamente atentando para esses e outros tipos de riscos potenciais, incidentes em maior ou menor grau a direitos juridicamente tutelados, que Peixoto (2020) aponta fatores críticos que não podem ser negligenciados para que se produza IA no Judiciário dentro de parâmetros éticos.

Com efeito, tendo em vista a dimensão do papel do Poder Judiciário no âmbito do Estado Democrático de Direito, como garantidor de direitos e liberdades fundamentais à sociedade, a questão dos riscos da utilização de ferramentas de IA pelos juízes adquire especial relevância. São enormes os desafios para cada juiz que, ao utilizar ferramentas de IA na função jurisdicional, necessita exercer efetiva supervisão daquelas, garantindo sua

aplicação ética. Conforme determina a Resolução nº 332/2020 do CNJ, o usuário – e, portanto, também o juiz-usuário – deve ter controle sobre a ferramenta, garantindo-se, desse modo, sua autonomia relativamente àquela. Trata-se de verdadeiro poder-dever, pois ao mesmo tempo em que busca garantir independência decisória ao magistrado, também lhe impõe o dever de supervisão da tecnologia para garantir à sociedade uma prestação jurisdicional legítima e justa.

O exercício do controle da IA pelo juiz-usuário, entretanto, pressupõe conhecimento para fazê-lo. Afinal, a tecnologia transforma a percepção humana e sua forma de vida (Ihde, 1979; Tripathi, 2017), exigindo dos juízes uma postura consciente e ativa na realização daquele.

Assim, a par do dever ético geral de competência atribuído aos magistrados, de acordo com os Princípios de Bangalore de Conduta Judicial (ONU, 2008), há um comando específico no sentido de informar-se para poder exercê-lo – afinal, os juízes devem ter consciência dos riscos de se acomodar a resultados automatizados, ou se deixar influenciar pela previsibilidade de reforma ou confirmação de suas decisões por instâncias recursais, deixando de cumprir seu dever de analisar todas as circunstâncias do caso concreto.

Considerando o panorama atual do desenvolvimento e da utilização de ferramentas de IA no Poder Judiciário brasileiro, o presente estudo, partindo do pressuposto de que o conhecimento é fundamental ao pleno exercício do controle pelo usuário-juiz, investiga se os juízes, usuários destas ferramentas, percebem-se ou não aptos a com elas operarem. Igualmente, investiga se os juízes brasileiros têm procurado formação específica quanto ao tema ou recebido oferta dessa formação das escolas judiciais dos tribunais respectivos.

Neste contexto, o presente artigo traça um panorama da IA no Poder Judiciário brasileiro e analisa os regramentos éticos estabelecidos pelo CNJ à luz da ética aplicada, inquirindo quanto à efetiva implementação das condições que permitam ao juiz-usuário exercer o poder-dever de controle que lhe incumbe.

1 UM PANORAMA DA INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO BRASILEIRO

O relatório Justiça em Números 2021, elaborado pelo Conselho Nacional de Justiça (2021, p. 11), informa que o Poder Judiciário brasileiro se compõe de 90 diferentes tribunais, além do Supremo Tribunal Federal.

São 27 Tribunais de Justiça Estaduais, 5 Tribunais Regionais Federais, 24 Tribunais Regionais do Trabalho, 27 Tribunais Regionais Eleitorais, 3 Tribunais de Justiça Militar Estaduais, o Superior Tribunal de Justiça, o Tribunal Superior do Trabalho, o Tribunal Superior Eleitoral e o Superior Tribunal Militar.

Segundo o relatório, durante o ano de 2020, a quase totalidade dos novos processos ingressou no Judiciário na forma digital, com apenas 3,1% na modalidade física (CNJ, 2021, p. 127), dado que bem demonstra o avançado nível de digitalização do processo judicial que se alcançou em todo o País.

Em paralelo, o relatório também aponta que o Poder Judiciário contava com um estoque de 75,4 milhões de processos pendentes no final do ano de 2020 (CNJ, 2021, p. 102), para uma força de trabalho de 22.695 magistrados – dos quais 20,7%, ou 4.707 cargos estão vagos (CNJ, 2021, p. 94). Quanto aos servidores, do total de 267.613, 224.001 são efetivos (83,7%) e, destes, há 49.662 cargos criados e não providos, representando 18% dos cargos (CNJ, 2021, p. 96).

Tais números indicam uma sobrecarga de trabalho humano que pode impactar diretamente no tempo de tramitação dos feitos, problema que pode ser amenizado de maneira significativa com a adoção de IA como ferramenta auxiliar para tarefas repetitivas e de grande volume, entre muitas outras possibilidades de aplicação.

Em dezembro de 2020, o Centro de Inovação, Administração e Pesquisa do Judiciário, da Fundação Getúlio Vargas (FGV) divulgou o relatório da primeira fase da pesquisa "Tecnologia aplicada à gestão dos conflitos no âmbito do Poder Judiciário com ênfase em inteligência artificial", coordenada pelo Ministro Luis Felipe Salomão, do Superior Tribunal de Justiça, que mapeou o desenvolvimento e a utilização de IA pelas Cortes de Justiça nacionais, verificando que até dezembro de 2020 já existiam ao menos 64 projetos de IA em 47 tribunais brasileiros, em variados estágios de desenvolvimento (FGV, 2020, p. 26).

Ainda de acordo com o documento, é possível visualizar que tais projetos começaram timidamente em 2018, expandindo-se em 2019 e dando um salto em 2020 (FGV, 2020, p. 67-68). Destaca-se, ademais, que os projetos de IA não acarretaram aumento significativo de despesas, porque se observou que "a série histórica de despesas com informática ficou praticamente estável" (FGV, 2020, p. 69).

O relatório da pesquisa informou que, ao tempo de sua publicação, em diferentes fases (desenvolvimento, projeto-piloto ou produção), havia 9 projetos de IA no Supremo Tribunal Federal, no Superior Tribunal de Justiça e no Tribunal Superior do Trabalho; 16 projetos distribuídos entre todos os 5 Tribunais Regionais Federais; 7 projetos em Tribunais Regionais do Trabalho e 31 projetos em Tribunais Estaduais (FGV, 2020, p. 66).

Quanto aos números de projetos apontados, fez-se a ressalva de que os dados foram coletados entre fevereiro e agosto de 2020, principalmente por formulário, que teve um retorno de 98%, e não 100%; e também que o número inicial de projetos era de 72, o que foi reduzido após a constatação de que alguns eram sistemas de TI, e não IA propriamente dita. Destacou-se ainda que, devido ao dinamismo no campo tecnológico, há uma necessidade constante de atualização dos números correspondentes (FGV, 2020, p. 8, 24 e 26).

A pesquisa da FGV também cuidou de verificar a origem do desenvolvedor das soluções de IA encontradas nos diversos tribunais, constatando que, da grande maioria que foi desenvolvida entre 2019 e 2020, 47 projetos foram elaborados internamente pelas próprias secretarias de tecnologia dos tribunais, 3 mediante parceria com universidades, 13 em parceria com uma empresa privada e 1 por outros órgãos (FGV, 2020, p. 69).

No tocante à aplicabilidade das ferramentas variadas de IA existentes nos tribunais, o relatório da pesquisa aponta (FGV, 2020, p. 69):

> De forma geral, os projetos de IA nos tribunais comportaram as seguintes funcionalidades: verificação das hipóteses de improcedência liminar do pedido nos moldes enumerados nos incisos do art. 332 do Código de Processo Civil; sugestão de minuta; agrupamento por similaridade; realização do juízo de admissibilidade dos recursos; classificação dos processos por assunto; tratamento de demandas de massa; penhora *on-line*; extração de dados de acórdãos; reconhecimento facial; *chatbot*; cálculo de probabilidade de reversão de decisões; classificação de petições; indicação de prescrição; padronização de documentos; transcrição de audiências; distribuição automatizada; e classificação de sentenças.

Por meio das funcionalidades apontadas, pode-se depreender que os tribunais estão buscando utilizar a IA principalmente para aumentar a produtividade, reduzir o tempo de tramitação dos processos, otimizar recursos humanos e materiais, e garantir mais segurança jurídica via respeito aos

precedentes – pela previsibilidade de aplicação das mesmas regras a casos semelhantes.

Em dezembro de 2020, o CNJ noticiou que passou a disponibilizar, em seu portal na Internet para consulta em tempo real, o Painel de Projetos com Inteligência Artificial, buscando conferir mais transparência a respeito de quantos tribunais estão utilizando a tecnologia e quais são as funcionalidades, destacando que 88% das iniciativas utilizam, em alguma medida, código-fonte na linguagem Python. O painel apresenta 41 projetos em 32 tribunais, contendo diversas informações e permitindo fazer *download* de relatórios (CNJ, 2020a).

Constata-se, portanto, que as ferramentas de IA estão avançando no Poder Judiciário brasileiro, com um número significativo de projetos em desenvolvimento ou já em plena utilização. Para os fins deste trabalho, resta investigar em que medida o conhecimento e a formação dos juízes usuários estão em compasso com o avanço tecnológico.

2 O PODER-DEVER ÉTICO DO JUIZ ENQUANTO USUÁRIO CONTROLADOR DA INTELIGÊNCIA ARTIFICIAL

Em dezembro de 2018, a Comissão Europeia para a Eficácia da Justiça (CEPEJ), órgão do Conselho da Europa, divulgou a Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu Ambiente, estabelecendo cinco princípios que devem nortear a aplicação de inteligência artificial no âmbito da Justiça: princípio do respeito dos direitos fundamentais; princípio da não discriminação; princípio da qualidade e da segurança; princípio da transparência, imparcialidade e equidade; princípio "sob controle do usuário".

No Brasil, o CNJ baixou a Resolução nº 332, de 21 de agosto de 2020, que "dispõe sobre a ética, a transparência e a governança na produção e no uso de inteligência artificial no Poder Judiciário", fazendo constar expressamente em seus "considerandos" que tomou por base a Carta Ética europeia. Naquela resolução, foram incorporados os cinco princípios mencionados *supra*, constantes da referida Carta.

Ademais, a Resolução nº 332/2020 atribuiu ao juiz, que se enquadra como usuário interno, o encargo de controlar a ferramenta de IA ao aplicá-la em sua atividade jurisdicional. Trata-se de tarefa deveras desafiadora, pois deve o magistrado ser capaz de detectar falhas nos dados utilizados ou

nos resultados apresentados pela IA, ou vieses revelando discriminação de gênero, raça e classe social, entre outras situações que podem ser prejudiciais aos jurisdicionados.

Além disso, como já mencionado, o juiz precisa ter consciência dos riscos de acomodação aos resultados automatizados oferecidos pela IA, e dos potenciais incentivos para tanto, particularmente em um contexto em que os magistrados ainda são majoritariamente avaliados pela produção numérica em comparação a outros critérios qualitativos de avaliação (CNJ, 2020b).

É interessante notar que a Resolução CNJ nº 332/2020 se preocupou em adotar providências concretas para mitigar o risco de vieses nos sistemas de IA usados na Justiça brasileira, a exemplo da busca por ampla diversidade expressa no art. 20 quanto às equipes de desenvolvedores, no que se refere a "gênero, raça, etnia, cor, orientação sexual, pessoas com deficiência, geração e demais características individuais". Ainda, de acordo com o § 1º do mesmo dispositivo, houve preocupação com que a participação representativa existisse em todas as etapas do processo (CNJ, 2020). A diversidade, ainda, deve ser observada para fins de distribuição de vagas de capacitação (CNJ, 2020).

Porém, muito embora tais providências sejam importantes e necessárias, não são suficientes em se tratando do magistrado. É preciso que o juiz, usuário-controlador, também tenha consciência dos possíveis riscos de discriminações algorítmicas, preparando-se para identificá-las e corrigi-las. Para tanto, capacitações em julgamento com perspectiva de gênero, em letramento racial e em proteção de dados são de fundamental importância, cabendo também aos Tribunais proverem essa necessidade.

Para o escopo deste trabalho, portanto, interessa uma análise mais detalhada das regras em torno da figura do usuário, enquanto controlador – por ser o papel que caberá ao juiz desempenhar. Nesse passo, no capítulo que trata das disposições gerais da Resolução nº 332/2020, há o conceito de usuário, desdobrado em interno e externo (CNJ, 2020).

Usuário é definido como a "pessoa que utiliza o sistema inteligente e que tem direito ao seu controle, conforme sua posição endógena ou exógena ao Poder Judiciário", conforme o art. 3º, IV, da Resolução. O juiz se enquadra na categoria de usuário interno, definido como "membro, servidor ou colaborador do Poder Judiciário que desenvolva ou utilize o sistema inteligente", nos termos do inciso V do citado dispositivo. Usuário externo,

por sua vez, é a "pessoa que, mesmo sem ser membro, servidor ou colaborador do Poder Judiciário, utiliza ou mantém qualquer espécie de contato com o sistema inteligente". São exemplos de usuários externos: advogados e jurisdicionados, entre outros atores constantes do rol exemplificativo no inciso VI do mesmo art. 3º da Resolução nº 332/2020 (CNJ, 2020).

A definição de usuário interno está relacionada ao fato de que o desenvolvimento de um sistema de IA é interdisciplinar, conforme o art. 20, *caput* e § 4º, da Resolução nº 332/2020, por ser feito em equipe e envolver tanto profissionais da área de Tecnologia da Informação (engenheiros e cientistas de dados, entre outros), que detêm o conhecimento técnico necessário à programação do sistema, como também profissionais da área em que o sistema inteligente será aplicado – no caso, a área jurídica –, pois é a partir das considerações, informações de ordem prática e necessidades concretas destes que os projetos são desenvolvidos (CNJ, 2020). No caso do Poder Judiciário, tais profissionais são principalmente os magistrados e servidores, prevendo-se também a participação de profissionais de outras áreas do conhecimento em conformidade com cada projeto específico.

Quanto aos usuários internos, incluindo o juiz, o art. 17, *caput* e inciso II, da Resolução nº 332/2020 estabelece que "o sistema inteligente deverá assegurar a autonomia dos usuários internos". Para tanto, deve possibilitar "a revisão da proposta de decisão e dos dados utilizados para sua elaboração, sem que haja qualquer espécie de vinculação à solução apresentada pela Inteligência Artificial" (CNJ, 2020).

De acordo com o parágrafo único do art. 18 da Resolução, destaca-se "o caráter não vinculante da proposta de solução apresentada pela inteligência artificial, a qual sempre é submetida à análise da autoridade competente" (CNJ, 2020).

Além disso, o art. 19, *caput* e parágrafo único, da Resolução nº 332/2020 dispõe que os sistemas de IA usados para a elaboração de decisão judicial, enquanto ferramenta de auxílio, "observarão, como critério preponderante para definir a técnica utilizada, a explicação dos passos que conduziram ao resultado", devendo sempre "permitir a supervisão do magistrado competente" (CNJ, 2020).

É possível notar, dessas disposições normativas, uma preocupação em reforçar o caráter auxiliar da IA no uso jurisdicional, vista como uma ferramenta sempre sujeita ao controle do juiz, o qual deverá ser capaz de aferir a observância dos parâmetros éticos da proposta elaborada pelo siste-

ma inteligente e corrigir erros, sabendo identificar se o erro partiu dos dados que alimentam o sistema ou da forma como a IA processou esses dados.

Ocorre que somente é possível controlar o que se conhece, o que remete a uma reflexão sobre a aplicação dos parâmetros éticos relativos ao juiz-controlador no âmbito do Poder Judiciário brasileiro.

3 EDUCAÇÃO E ÉTICA PARA USO DE INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO

O debate que se inicia a partir da definição de parâmetros éticos para o exercício adequado da função jurisdicional com o uso da IA remete, necessariamente, para a discussão do conceito de ética e do que ela enseja do ponto de vista prático. O que poderia ser definido como ética na Resolução nº 332/2020? Certamente, a leitura da referida Resolução conduz a uma série de reflexões que não podem ser esgotadas em um único artigo. Fica claro, no entanto, que a IA deve estar situada no "lugar" de mera auxiliar no cenário jurídico, de modo que não haja risco que se firam os direitos fundamentais, bem como sejam garantidos e preservados a transparência, a segurança e o controle do usuário.

Chama atenção, contudo, o fato de já existirem inúmeros projetos de IA em andamento sem que exista uma política de formação instituída para que tais iniciativas aconteçam de forma relativamente segura e em consonância com o título da própria "Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu Ambiente", que inspirou o ato normativo brasileiro. Os cinco princípios que versam sobre "Direitos Fundamentais; Não Discriminação; Qualidade e Segurança; Transparência, Imparcialidade e Equidade; Sob Controle do Usuário" parecem garantir que a questão ética seja valorizada e, ao mesmo tempo, conduzem o cidadão a uma conclusão básica de que, sem formação educativa para o uso da IA no contexto judiciário, tal perspectiva passa a gozar de perigosa fragilidade quando a meta a ser alcançada é o bem público. Nessa esteira, parece fundamental primeiramente revisitar o conceito de ética e, por conseguinte, examinar de que maneira o Judiciário pode valer-se dela para utilizar recursos de IA em seu cotidiano.

De acordo com Cortina e Martinez (2005), a ética não tem como função apenas o esclarecimento e a fundamentação do fenômeno da moralidade, mas também a aplicação de suas descobertas aos diferentes âmbitos da vida social. Em conjunto com a missão de fundamentação está a tarefa da aplicação, que consiste em averiguar como os princípios ajudam a orientar

os diferentes tipos de atividade. Os autores ressaltam que não basta refletir sobre como se aplicam os princípios em cada âmbito concreto, fazendo-se necessário levar em conta a especificidade de cada atividade com suas próprias exigências morais e seus próprios valores. Trata-se, pois, de verificar quais são os bens internos que cada tipo de atividade deve trazer para a sociedade e quais são os valores e hábitos que é preciso incorporar para alcançá-los. Para chegar a isso, deve-se trabalhar interdisciplinarmente junto aos especialistas de cada área, por meio da implementação de uma cultura educativa comprometida e constante que seja capaz de ampliar o olhar ético para o moral cívico, regente do tipo de sociedade em que vivemos.

O uso ético de recursos de IA no Judiciário carece, ao menos, da organização de uma política formativa contundente para que os usuários internos estejam minimamente esclarecidos a respeito do funcionamento e/ou impacto que esse tipo de tecnologia pode provocar. O viés prático da questão ética está, nesse caso, inteiramente vinculado a uma proposição educacional consistente que priorize a aprendizagem dos magistrados e servidores, capacitando-os para serem também agentes éticos quando da utilização desse tipo de recurso no âmbito jurisdicional. Nesse caso, a ética só se faz prática mediante a existência de um processo de ensino que permita a adequada preparação do magistrado para o exercício consciente e esclarecido da sua função quando auxiliado por esse tipo de recurso. Importante esclarecer que, caso a questão ética se encerre apenas na deflagração de princípios sem que haja uma preocupação com a sua tradução concreta, o ferimento à própria ética como princípio acaba por se constituir.

Os filósofos alemães da Escola de Frankfurt, durante o século XX, enfatizaram a importância de que a postura geral da sociedade fosse pautada pela ética, pelo cuidado com o outro, pela alteridade e pela solidariedade, ressaltando o dever humano de repensar constantemente as situações vividas e reconhecer o sofrimento alheio. Nesse sentido, tal forma social de ser teria a importante função de fortalecer seu vínculo com a ética e de conduzir os indivíduos à maior amplitude de experiências. Seguindo essa orientação, a relação entre ética e educação deve ser predominantemente refletida para que possa ser incorporada ao campo das ações. Conforme Adorno (1995), a educação não se restringe exclusivamente à pauta do desenvolvimento das estratégias de esclarecimento da consciência, mas tem o compromisso de levar em conta e em grande medida a forma social em que ela se dá, concretizando-se, também, como apropriação de conhecimentos técnicos. Portanto, compreender que o atendimento ao compromisso ético do Judiciário frente à implementação de projetos de IA passa necessariamente por

um vultoso dever educacional é condição sine qua non para que a Carta Europeia e a Resolução nº 332/2020 possam justificar a sua existência.

Veja-se que o princípio da competência se inclui entre os Princípios de Bangalore de Conduta Judicial, de acordo com o art. 6.3. Tal princípio se desdobra em deveres tanto para o juiz – que deve buscar manter e aprimorar seu conhecimento, suas habilidades e suas qualidades pessoais necessárias ao desempenho adequado de suas funções – quanto para os tribunais – que devem proporcionar treinamento adequado aos juízes (ONU, 2008).

Ademais, o CNJ recentemente promoveu alterações na Resolução nº 75/2009, que trata dos concursos públicos para a magistratura, exigindo, entre outros conteúdos, conhecimentos sobre inteligência artificial e outras tecnologias, em deliberação unânime do Ato Normativo nº 0006767-49.2021.2.00.000, por ocasião da 93ª Sessão Virtual do CNJ, encerrada em 24 de setembro de 2021 (CNJ, 2021b).

Assim, quanto aos magistrados que já se encontram na carreira, cabível perquirir se a utilização e o desenvolvimento de sistemas de IA nos tribunais vêm sendo acompanhados do dever ético educativo correspondente, indispensável a que o juiz possa efetivamente exercer o controle enquanto usuário de ferramentas auxiliares de IA.

4 ANÁLISE DA PESQUISA EMPÍRICA

O presente estudo, como antes mencionado, investiga a percepção dos juízes brasileiros quanto a seus conhecimentos sobre IA, bem como sua busca por capacitação e a oferta disponibilizada pelas escolas judiciais. Para tanto, a pesquisa se utilizou de uma metodologia quantitativa, de natureza aplicada, com objetivos exploratórios e procedimentos do tipo *survey*.

Como instrumento de coleta de dados, estruturou-se um questionário com termo de consentimento livre e esclarecido contendo 8 perguntas, que foi respondido por 688 magistrados integrantes de 4 dos 5 ramos da Justiça brasileira, a saber: federal, estadual, trabalhista e militar. Optou-se por não incluir o ramo da justiça eleitoral, visto que tal jurisdição é exercida por juízes estaduais, nos termos do art. 121 da Constituição Federal (Brasil, 1988), para evitar duplicidade de respostas.

No universo de 17.988 cargos providos de magistrados, de acordo com o relatório Justiça em Números 2021 (CNJ, 2021), o quantitativo de 688 respostas obtidas entre os dias 14 e 21 de setembro de 2021 ultrapassa

a amostra ideal de 642 respondentes, permitindo atingir 99% de nível de confiança, com 5% de margem de erro, de acordo com a seguinte fórmula:

Figura 1: Fórmula estatística utilizada na pesquisa.

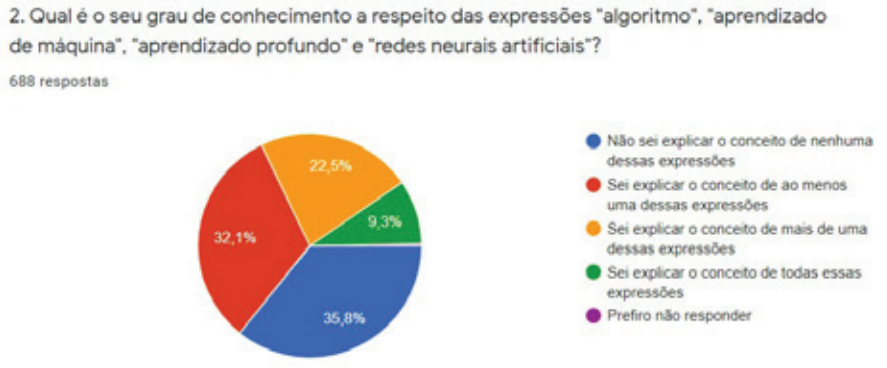
$$n = \frac{k^2 qpN}{e^2(N-1) + k^2 pq}$$

Fonte: Questionpro, 2021.

Em que (N) é o universo total de juízes que poderiam responder à pesquisa; (e) é a margem de erro; (k) é o nível de confiança; (p) é a probabilidade de sucesso; e (q) é a probabilidade de fracasso.

Quando perguntados sobre seu grau de conhecimento a respeito das expressões "algoritmo", "aprendizado de máquina", "aprendizado profundo" e "redes neurais artificiais", apenas 9,3% dos respondentes disseram saber explicar o conceito de todas essas expressões. Como se pode ver do gráfico a seguir (questão 2), 35,8% dos respondentes não sabiam explicar o conceito de nenhuma, enquanto os demais (54,6%) reportaram saber explicar o conceito de apenas parte das expressões:

Gráfico 1: Respostas ao segundo questionamento.



Fonte: Elaborado pelas autoras.

Os juízes reportaram, em sua esmagadora maioria, ter pouco ou nenhum conhecimento sobre a Resolução nº 332/2020 do CNJ, o que indica uma possível falta de conhecimento também de seu papel enquanto con-

troladores na utilização de ferramentas de IA – papel imposto pela referida Resolução (a seguir, questão 3):

Gráfico 2: Respostas ao terceiro questionamento.

Fonte: Elaborado pelas autoras.

Ainda, 61,3% dos juízes relataram se sentir totalmente despreparados para exercer o controle e a supervisão de sistemas auxiliares de IA para a elaboração de decisões judiciais, sendo que apenas 8,7% se consideraram bem preparados para tanto (questão 4, a seguir):

Gráfico 3: Respostas ao quarto questionamento.

Fonte: Elaborado pelas autoras.

Perguntados sobre o quão preparados se sentiam para julgar litígios envolvendo a aplicação de IA, 57,3% afirmaram sentir-se totalmente despreparados, e apenas 9,3% reportaram estar bem preparados para tanto (questão 5, a seguir):

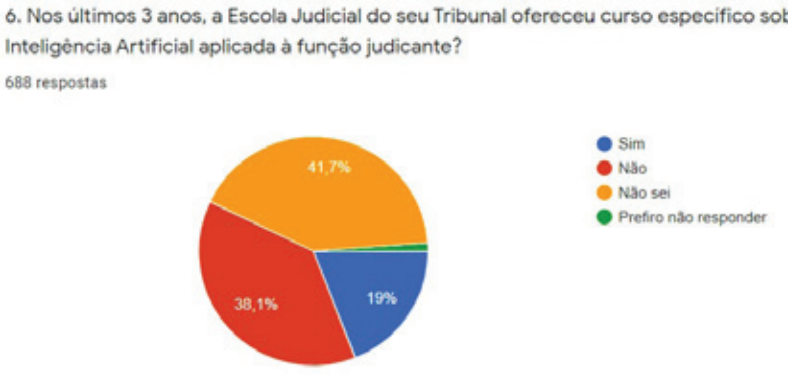
Gráfico 4: Respostas ao quinto questionamento.

Fonte: Elaborado pelas autoras.

Veja-se que, pelo gráfico *supra*, a maioria dos juízes não apenas se considera totalmente despreparada para exercer *o controle e a supervisão* de sistemas auxiliares de IA para a elaboração de decisões judiciais, como também se considera totalmente despreparada para *julgar litígios* envolvendo a aplicação de IA.

Quanto à capacitação, embora 38,1% dos respondentes tenham afirmado que as escolas judiciais de seus tribunais respectivos não ofereceram nenhum curso específico sobre IA aplicada à função judicante, chama a atenção que 41,7% *desconhecem* se a respectiva escola judicial ofereceu ou não tal formação nos últimos 3 anos (a seguir, questão 6):

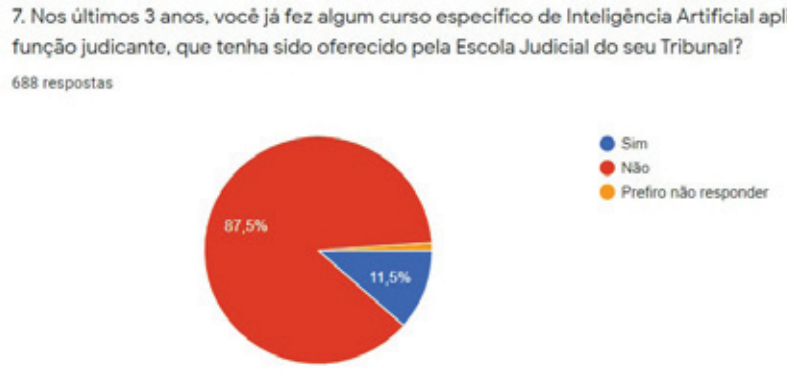
Gráfico 5: Respostas ao sexto questionamento.



Fonte: Elaborado pelas autoras.

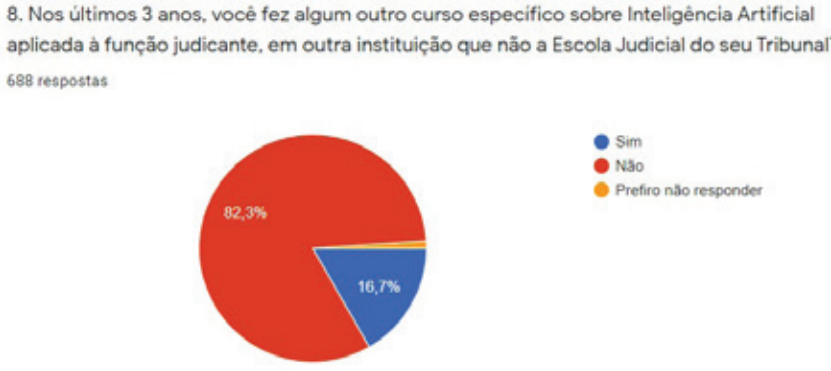
Não obstante a percepção de despreparo para trabalhar com IA, os magistrados não cursaram formações específicas em IA aplicada à função judicante nos últimos três anos, nem nas escolas judiciais dos tribunais respectivos – 87,5% (questão 7, a seguir) – nem em qualquer outra instituição – 82,3% (questão 8, a seguir):

Gráfico 6: Respostas ao sétimo questionamento.



Fonte: Elaborado pelas autoras.

Gráfico 7: Respostas ao oitavo questionamento.



Fonte: Elaborado pelas autoras.

Analisada em seu conjunto, a pesquisa revela que os magistrados brasileiros pouco ou nada conhecem sobre a Resolução nº 332/2020 do CNJ, de onde se pode validamente depreender que tampouco estão devidamente informados sobre seu poder-dever enquanto usuários controladores das ferramentas de IA com que venham a trabalhar, ou já estejam trabalhando, em suas funções jurisdicionais.

Tal desconhecimento pode explicar, em alguma medida, o fato de que, apesar de os juízes se perceberem como despreparados para exercer o controle e a supervisão de ferramentas de IA, não terem ainda buscado formação específica para suprir essa deficiência.

Ademais, a pesquisa indica também a falta de oferta – ou de devida comunicação das ofertas existentes – de cursos específicos por parte das escolas judiciais, sendo possível – hipótese que, naturalmente, dependeria de investigação mais aprofundada – que também as administrações dos diversos tribunais e das escolas judiciais não estejam adequadamente informadas sobre os deveres impostos pela Resolução nº 332/2020 do CNJ.

Os dados coletados, portanto, quando confrontados com o fato de que há ao menos 64 projetos de IA nos tribunais brasileiros, inclusive levando à edição de uma resolução específica para tratar da matéria pelo CNJ, sugerem que: a) os investimentos maciços no desenvolvimento e na adoção de ferramentas de IA pelos tribunais não têm sido distribuídos de forma a contemplar a capacitação necessária para que os juízes exerçam o papel de

usuário-controlador que deles se espera; b) a Resolução nº 332/2020, muito embora louvável enquanto normativo para regular o desenvolvimento e o uso de IA pelos tribunais brasileiros, não vem sendo comunicada de forma eficiente aos magistrados, em especial chamando a atenção para seu dever enquanto usuários-controladores; c) as escolas judiciais tampouco se mobilizaram para capacitar os juízes a cumprir com o poder-dever que lhes incumbe por força da Resolução nº 332/2020, o que também poderia se explicar por uma comunicação ineficiente daquela aos tribunais e pelo próprio fato de que aquele normativo não incluiu entre seus dispositivos o dever de capacitação; d) os juízes que vêm utilizando ferramentas de IA nos diversos tribunais não estão plenamente capacitados para exercer a função de usuários controladores, em prejuízo ao cumprimento do princípio ético "do controle do usuário", previsto no Capítulo VII da Resolução nº 332/2020.

A pesquisa, portanto, acende um sinal de alerta quanto ao uso ético de IA pelos magistrados brasileiros. Ela aponta para a necessidade de reflexão sobre a atual estratégia de investimentos em uso e desenvolvimento dessa tecnologia, para também incluir ou ampliar investimentos na capacitação dos magistrados e na comunicação do que deles se espera enquanto usuários-controladores, na forma da Resolução nº 332/2020. Veja-se que a provável e rápida sofisticação dos sistemas recomenda a pronta atuação do Poder Judiciário nesse sentido, a partir de uma estratégia eficiente de comunicação que engaje magistrados e escolas judiciais na tarefa de promover e buscar aprendizagem.

CONSIDERAÇÕES FINAIS

Os novos desafios impõem ao juiz contemporâneo ser mais tecnológico e ao mesmo tempo mais humano, buscando o aperfeiçoamento contínuo, não só para exercer suas funções em um Judiciário cada vez mais digital, mas para se manter atento à evolução da sociedade e ao dever de garantir a todos o direito fundamental de acesso à justiça.

Quando se analisa o cenário atual de desenvolvimento de projetos sobre IA aplicada ao Poder Judiciário, percebe-se um ritmo intenso e acelerado, com foco, sobretudo, no aumento da produtividade e na redução do tempo de tramitação dos processos. Contudo, embora seja inegável o potencial de contribuição da IA para atingir o objetivo maior da melhoria da prestação do serviço jurisdicional, não se pode ter em mente apenas

aspectos quantitativos: é fundamental atentar também para a questão da qualidade.

A independência do Judiciário confere direitos e prerrogativas ao juiz, constitucionalmente assegurados, mas também lhe impõe obrigações éticas. Entre elas, inclui-se o dever de executar o trabalho judicial com competência e diligência. Isso implica dizer que o juiz deve ter substancial habilidade profissional, adquirida, mantida e regularmente reforçada por treinamento ao qual ele tem não apenas o dever, mas também o direito, de submeter-se.

Morley e colaboradores (2020, p. 2147) defendem que é chegada a hora de se passar para a "segunda fase" da ética em IA, aquela que traduz entre "o quê" para "como". Segundo eles: "[...] the gap between principles and practice is large, and widened by complexity, variability, subjectivity, and lack of standardization, including variable interpretation of the 'components' of each of the ethical principles"[4]. Fazendo um paralelo com a problemática trazida no presente artigo, não basta normatizar o poder-dever do juiz enquanto usuário-controlador da IA para garantir sua utilização ética pelo Poder Judiciário: é preciso comunicar amplamente esse poder-dever e garantir a capacitação necessária para que ele seja devidamente exercido.

Segundo o sumário executivo divulgado em dezembro de 2019 sobre o "Estudo da Imagem do Poder Judiciário brasileiro", encomendado pela Associação dos Magistrados Brasileiros (AMB) à Fundação Getúlio Vargas (FGV), com a participação do Instituto de Pesquisas Sociais, Políticas e Econômicas (Ipespe), os atributos mais importantes que representariam um sistema de justiça ideal, na visão da sociedade, são um Judiciário confiável (41%), imparcial/igual para todos (35%) e transparente (34%).

A pesquisa de que trata este artigo sugere que o uso da IA pelos tribunais brasileiros não vem sendo acompanhado do pleno exercício do controle pelo juiz-usuário, uma vez que os juízes não apenas se sentem despreparados para realizar tal controle, como também desconhecem essa incumbência que lhes é atribuída. Ademais, a pesquisa também revela a carência de um movimento consistente por parte das escolas judiciais, no sentido de promover a ampla capacitação com foco nos juízes brasileiros, preparando-os para essa intensa revolução tecnológica hoje vivenciada. À

4 "O vão entre princípios e prática é grande, e ampliado pela complexidade, variabilidade, subjetividade, e falta de padronização, incluindo interpretação variável dos 'componentes' de cada um dos princípios éticos." (Tradução livre de Morley et al., 2020, p. 2147).

míngua de medidas para contornar tais deficiências, corre-se o risco de produzir impactos indesejáveis na atividade jurisdicional, deixando o Judiciário de corresponder aos atributos dele esperados pela sociedade.

A existência de diversos sistemas de IA já em uso pelo Judiciário convida, portanto, à reflexão quanto ao pleno atendimento dos princípios de uma IA ética por parte da instituição.

Assim, o presente artigo alerta para a necessidade de que o Poder Judiciário brasileiro amplie a estratégia atualmente aplicada ao uso e ao desenvolvimento de sistemas de IA para incorporar, de modo mais efetivo, as diretrizes éticas correspondentes – em especial, o princípio "sob controle do usuário". Consequentemente, propõe que, ao lado dos investimentos no desenvolvimento técnico das ferramentas, também se invista em comunicação e na capacitação dos magistrados para que desenvolvam as competências necessárias para garantir uma IA ética no Poder Judiciário brasileiro.

REFERÊNCIAS

ADORNO, Theodor W. *Educação e emancipação*. Rio de Janeiro: Paz e Terra, 1995.

Brasil. Constituição da República Federativa do Brasil de 1988, de 5 de outubro de 1988. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 28 set. 2021.

COMISSÃO EUROPEIA PARA A EFICÁCIA DA JUSTIÇA (CEPEJ). Carta Europeia de Ética sobre o Uso da Inteligência Artificial em Sistemas Judiciais e seu ambiente. Disponível em: https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0. Acesso em: 28 set. 2021.

CONSELHO NACIONAL DE JUSTIÇA (CNJ). Painel dá transparência a projetos de inteligência artificial no Judiciário. Notícias CNJ/Agência CNJ de Notícias, 23 dez. 2020. Disponível em: https://www.cnj.jus.br/painel-da-transparencia-a-projetos-de-inteligencia-artificial-no-judiciario/. Acesso em: 29 set. 2021.

_____. Justiça em Números 2021. Brasília: CNJ, 2021. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2021/09/relatorio-justica-em-numeros2021-12.pdf. Acesso em: 29 set. 2021.

_____. Resolução nº 75, de 12 de maio de 2009. Disponível em: https://atos.cnj.jus.br/atos/detalhar/100. Acesso em: 28 set. 2021.

_____. Resolução nº 332, de 21 de agosto de 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em: 28 set. 2021.

_____. Resolução nº 325, de 29 de junho de 2020. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3365. Acesso em: 28 set. 2021.

CORTINA, Adela; MARTINEZ, Emílio. *Ética*. São Paulo: Loyola, 2005.

FUNDAÇÃO GETÚLIO VARGAS (FGV). Estudo da imagem do Judiciário brasileiro. AMB/FGV/Ipesp, 2019. Disponível em: https://www.amb.com.br/wp-content/uploads/2020/04/ESTUDO_DA_IMAGEM_DO_JUDICIARIO_BRASILEIRO_COMPLETO.pdf. Acesso em: 28 set. 2021.

_____. Inteligência artificial: tecnologia aplicada à gestão de conflitos no âmbito do Poder Judiciário. Coordenação Luis Felipe Salomão, 2020. Disponível em: https://ciapj.fgv.br/sites/ciapj.fgv.br/files/estudos_e_pesquisas_ia_1afase.pdf. Acesso em: 28 set. 2021.

IHDE, Don. *Technics and Praxis*. Dordrecht: D. Reidel Publishing Company, 1979.

KEHL, Danielle Leah; KESSLER, Samuel Ari. Algorithms in the criminal justice system: Assessing the use of risk assessments in sentencing, 2017. Disponível em: https://dash.harvard.edu/handle/1/33746041. Acesso em: 28 set. 2021.

MORLEY, Jessica et al. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. *Science and Engineering Ethics*, v. 26, n. 4, p. 2141-2168, 2020. Disponível em: https://link.springer.com/article/10.1007/s11948-019-00165-5%23Sec2. Acesso em: 28 set. 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Comentários aos Princípios de Bangalore de Conduta Judicial/Escritório contra Drogas e Crime. Trad. Marlon da Silva Malha e Ariane Emílio Kloth. Brasília: Conselho da Justiça Federal, 2008. Tradução de: Commentary on The Bangalore Principles of Judicial Conduct. Disponível em: https://www.unodc.org/documents/lpo-brazil/Topics_corruption/Publicacoes/2008_Comentarios_aos_Principios_de_Bangalore.pdf. Acesso em: 28 set. 2021.

PEIXOTO, Fabiano Hartmann. *Inteligência artificial e direito*: convergência ética e estratégica. Coleção Direito, Racionalidade e Inteligência Artificial. 1. ed. Curitiba: Alteridade, 2020. 170 p.

QUESTIONPRO. Calculadora de amostra de pesquisa, 2021. Disponível em: https://www.questionpro.com/pt-br/mobile-diaries.html. Acesso em: 28 set. 2021.

TRIPATHI, Arun Kumar. Hermeneutics of technological culture, 2017. Disponível em: https://link.springer.com/content/pdf/10.1007/s00146-017-0717-4.pdf. Acesso em: 28 set. 2021.

Sobre as autoras:

Eunice M. B. Prado | *E-mail:* euniceprado_@hotmail.com

Juíza de Direito do Tribunal de Justiça de Pernambuco. Mestranda no Curso de Mestrado Profissional em Direito do PPGPD/Enfam – Escola Nacional de Formação e Aperfeiçoamento de Magistrados. Pós-Graduada em Direito do Consumidor e Responsabilidade Civil. Integrou no CNJ o Grupo de Trabalho "Ética e Inteligência Artificial", cujos estudos colaboraram para a Resolução nº 332/2020.

Luciane A. Corrêa Münch | *E-mail:* lucianemunch@icloud.com

Desembargadora Federal do Tribunal Regional Federal da 4ª Região. Membro do Corpo Docente do PPGPD/Enfam – Escola Nacional de Formação e Aperfeiçoamento de Magistrados. Doutora em Direito pela Universidade de Genebra, Suíça. LLM Master of Laws pela LSE Law School, Reino Unido. Mestre em Direito pela PUCRS. Especialista em Direito Internacional e Bacharel em Ciências Jurídicas e Sociais pela UFRGS.

Márcia A. Corrêa Ughini Villarroel | *E-mail:* marcia.correa@sertao.ifrs.edu.br

Doutora em Psicologia e Educação pela Universidade de São Paulo (USP). Professora Permanente do Mestrado em Informática da Educação do Instituto Federal do Rio Grande do Sul (IFRS), *Campus* Porto Alegre.

Data de Submissão: 30 de setembro de 2021.
Data de Aceite: 10 de janeiro de 2022.

Da "Caixa-Preta" à "Caixa de Vidro": o Uso da *Explainable Artificial Intelligence* (XAI) para Reduzir a Opacidade e Enfrentar o Enviesamento em Modelos Algorítmicos

From "Black Box" to "Glass Box": using Explainable Artificial Intelligence (XAI) to Reduce Opacity and Address Bias in Algorithmic Models

MARCO ANTÔNIO SOUSA ALVES[1]
Universidade Federal de Minas Gerais (UFMG).

OTÁVIO MORATO DE ANDRADE[2]
Universidade Federal de Minas Gerais (UFMG).

RESUMO: A inteligência artificial (IA) tem sido utilizada em larga escala em variados domínios, com cada vez mais implicações sociais, éticas e de privacidade. À medida que suas potencialidades e aplicações são expandidas, surgem dúvidas sobre a confiabilidade dos sistemas equipados com IA, particularmente aqueles que empregam técnicas de *deep learning* que podem torná-los verdadeiras "caixas-pretas". A XAI (*explainable artificial intelligence*), ou inteligência artificial explicável, objetiva oferecer informações que ajudam a explicar o processo preditivo de determinado modelo algorítmico. Este artigo se volta especificamente para o estudo da XAI, investigando seu potencial para explicar decisões de modelos algorítmicos e combater o enviesamento dos sistemas de IA. Na primeira etapa do trabalho, é discutida a questão da falibilidade e enviesamento da IA, e como a opacidade agrava esses problemas. Na segunda parte, apresenta-se a inteligência artificial explicável e suas potenciais contribuições para tornar os sistemas mais transparentes, auxiliando no combate aos erros e vieses algorítmicos. Conclui-se que a XAI pode colaborar para a identificação de vieses em modelos algorítmicos, razão pela qual se sugere que a capacidade de "se explicar" – ou seja, a *explicabilidade* – seja um requisito para a adoção de sistemas de IA em searas mais sensíveis, como, por exemplo, o auxílio à tomada de decisão judicial.

PALAVRAS-CHAVE: XAI; inteligência artificial explicável; opacidade algorítmica; transparência.

1 Orcid: http://orcid.org/0000-0002-4885-8773.
2 Orcid: http://orcid.org/0000-0002-0541-7353.

ABSTRACT: Artificial intelligence (AI) has been used on a large scale in different domains, with increasing social, ethical and privacy implications. As their potential and applications expand, concerns arise about the reliability of AI systems, particularly those that use deep learning techniques that can make them true "black boxes". XAI (explainable artificial intelligence) aims to offer information that helps explain the predictive process of a given algorithmic model. This article focuses specifically on the study of XAI, investigating its potential to explain algorithmic model decisions and combat bias in AI systems. In the first stage of the work, the issue of AI fallibility and bias is discussed – and how opacity intensifies these problems. The second part presents explainable artificial intelligence and its potential contributions to make systems more transparent, helping to combat algorithmic errors and biases. It is concluded that XAI can contribute to the identification of biases in algorithmic models, then it is suggested that the ability to "explain" should be a requirement for the adoption of AI systems in sensitive areas, such as help in court decisions.

KEYWORDS: XAI; explainable artificial intelligence; algorithmic opacity; transparency.

INTRODUÇÃO

Nos últimos anos, sistemas dotados de inteligência artificial (IA) têm invadido nossas vidas e otimizado processos nos mais variados campos, como, por exemplo, nas recomendações de buscas, produtos e preferências do usuário, no atendimento automático feito por *chatbots*, em diagnósticos médicos e em dispositivos "inteligentes" (*smart devices*), como os carros autômatos.

O avanço e a popularização da IA suscitam extensa gama de preocupações, desde aspectos relacionados à privacidade – vulnerável frente ao poder preditivo das *big techs* – passando pelo comprometimento de traços característicos da subjetividade humana (Rouvroy; Berns, 2015), até implicações éticas sobre vieses discriminatórios. Também se trava efervescente debate sobre como novos fenômenos relacionados à IA poderiam afetar a democracia e o pluralismo político, por meio da disseminação de *fake news*, radicalização ideológica ou mesmo da sofisticação de técnicas de censura e vigilância em massa (Sunstein, 2009; Bruno, 2013).

Tais preocupações são intensificadas pelo fato de que o funcionamento interno de determinados algoritmos – em particular aqueles dotados de aprendizado de máquina profundo (*deep learning*[3]) – pode ser um mistério completo para o usuário médio de tecnologia e, não raro, até para aqueles com competências avançadas na área. Enquanto no aprendizado de máquina (*machine learning*[4]) existe uma estrutura de aprendizado estatístico mais enxuta entre a entrada de dados (*input*) e a saída (*output*), no caso do *deep learning*, existem múltiplas camadas de redes neurais que se sobrepõem umas às outras, tornando mais complexa a compreensão do seu raciocínio. Desta forma, quando se trata de sistemas envolvendo múltiplas redes neurais artificiais, até o momento são escassas as ferramentas e técnicas capazes de facilitar o entendimento das decisões tomadas por um algoritmo (Cortiz, 2021).

É preciso considerar que a IA não é uma tecnologia em si, mas sim uma área do conhecimento, formada por diferentes vertentes, incluindo a *machine learning*, que não deve ser vista como sinônimo de "caixa preta". Afinal, nem toda técnica baseada em aprendizado de máquina padece de opacidade algorítmica. As técnicas de *machine learning* supervisionadas, por exemplo, são facilmente explicáveis. Neste artigo, nosso foco será direcionado para as técnicas não supervisionadas, especialmente o *deep learning* ou a aprendizagem profunda, que pode ser compreendida como uma subárea do aprendizado estatístico ou *machine learning*. Nesse domínio, a questão da explicação e a da regulação assumem um contorno bem mais delicado e complexo.

Diante dessa "opacidade algorítmica", ou seja, a incapacidade de se enxergar além do *output* produzido, questiona-se se o ser humano deveria delegar decisões tão importantes a sistemas de IA, nos casos em que estes são incapazes de explicar como chegaram a algumas conclusões. Com efeito, muitos pesquisadores que se debruçam sobre assunto têm considerado que é fundamental equipar os sistemas de IA com funcionalidades capazes de fornecer uma explicação razoável sobre suas predições

3 *Deep learning* é um método de aprendizado de máquina que usa redes neurais artificiais com várias camadas intermediárias entre a camada de entrada (*input*) e a camada de saída (*output*) e, portanto, uma extensa estrutura interna. Esse modelo tem a capacidade de ampliar suas camadas de redes neurais para solucionar o problema enfrentado (Cortiz, 2021).

4 *Machine learning* é um processo no qual um sistema artificial utiliza métodos estatísticos para aprender a partir de exemplos. Por apresentarem uma estrutura mais simples, os algoritmos de *machine learning* tendem a ser mais passíveis de entendimento do que os algoritmos que usam aprendizagem profunda (Cortiz, 2021).

(Villani, 2019, p. 114; Confalonieri *et al.*, 2020). Dentre as opções apresentadas como aptas a fornecer tal explicação, figuram os sistemas que têm sido chamados de "inteligência artificial explicável" (*explainable artificial intelligence – XAI*).

O presente trabalho tem por objetivo investigar, especificamente, a XAI (*explainable artificial intelligence*) como forma de reduzir a opacidade dos modelos algorítmicos. Estudos sobre o tema têm sugerido que sistemas habilitados para XAI, ao reduzir a opacidade, auxiliam na revelação de falhas nos algoritmos, proporcionando a oportunidade de se corrigir, ou ao menos minimizar, o enviesamento de máquina, tornando estes sistemas mais confiáveis (Ribeiro *et al.*, 2016; Nunes; Andrade, 2021; Wells; Bednarz, 2021).

Gostaríamos de deixar claro, contudo, que não pretendemos com este trabalho defender uma solução puramente tecnológica para todos os problemas éticos e políticos trazidos pelo uso de sistemas dotados de inteligência artificial. Está longe de nossa intenção abraçar uma posição tecnófila ingênua, que aposta simplesmente na possibilidade de a tecnologia resolver por si mesma todos os problemas que ela mesmo suscita, por meio apenas de sistemas cada vez mais sofisticados. Precisamos estar atentos aos limites de tal empreitada, identificando as áreas mais sensíveis e fixando balizas claras quanto à forma e, também, quanto à possibilidade mesma do uso da inteligência artificial. Apesar disso, entendemos que é possível aprimorar os sistemas existentes e ampliar sua possibilidade de uso adequado e responsável em diversos domínios, por meio, por exemplo, de mecanismos mais confiáveis e transparentes.

Este trabalho divide-se em duas grandes seções. Primeiramente, é problematizada a questão da falibilidade e dos vieses nos modelos algorítmicos, discorrendo-se sobre alguns casos em que os sistemas de IA forneceram resultados de predição (*outputs*) imprecisos ou equivocados, ou que, mesmo quando corretos, empreenderam um raciocínio que desconsiderou requisitos desejáveis. Na segunda etapa do trabalho, são apresentadas noções fundamentais sobre a XAI, explicitando-se algumas técnicas que atuam para reduzir a opacidade dos sistemas de IA. Ao final desta etapa, são retomados alguns exemplos de falibilidade e vieses elencados na primeira seção, analisando-se como técnicas correlatas de inteligência artificial explicável poderiam endereçar soluções em cada um desses casos. O trabalho conclui que a XAI pode contribuir para a identificação e o tratamento de problemas em modelos algorítmicos, razão pela qual se sugere que a capa-

cidade de "se explicar" – ou seja, a *explicabilidade* – seja um requisito para a adoção de sistemas de IA em searas mais sensíveis, como, por exemplo, o auxílio à tomada de decisão judicial.

1 QUANDO A IA FRACASSA: O PROBLEMA DE FRACASSAR NO ESCURO

1.1 Falibilidade da IA

Por mais sofisticado que seja, um sistema de IA não está isento de produzir resultados imprecisos, incompletos ou tendenciosos. Várias razões podem estar por trás de um resultado preditivo errôneo. Antes de tudo, os dados de entrada (*input*) podem ser inacabados ou conflitantes entre si, gerando ambiguidades para o algoritmo[5] que os analisa. Além disso, a predição computacional pode estar mal calibrada ou insuficientemente treinada, falhando na interpretação desses dados e terminando por fornecer resultados incorretos (Ramos, 2020). Por fim, há ainda os casos em que o algoritmo "acerta" a resposta, lançando mão, no entanto, de raciocínios e aproximações não desejáveis.

A *falibilidade* ocorre quando um sistema fracassa ao correlacionar os dados de modo causal, gerando evidências inconclusivas e ações injustificadas (Rossetti; Angeluci, 2021, p. 8). Um exemplo de falibilidade é a confusão feita por algoritmos classificadores de imagens ao tentarem distinguir entre lobos e cães da raça Husky, principalmente quando há presença de neve na imagem. Como a maioria das fotos de lobos nos dados de treinamento continha um fundo de neve, o algoritmo acabava se pautando pelo ambiente – e não pelas características do animal – para classificar a imagem. Ao analisar um cão Husky retratado contra um fundo de neve, o sistema falhava, classificando-o incorretamente como "lobo" (Ribeiro *et al.*, 2016).

5 Por algoritmo entende-se "um recurso crescente de agência de autoaprendizagem, interativa, autônoma, que permite que artefatos computacionais executem tarefas que, de outra forma, exigiriam que a inteligência humana fosse executada com sucesso" (Floridi; Taddeo, 2018, p. 751 – tradução nossa).

Figura 1: Um cão Husky (à esquerda) é confundido com um lobo, porque o fundo nevado (à direita) era relacionado erroneamente aos lobos. Essa falha se deve a dados de treinamento insuficientes (muitos lobos fotografados contra a neve), o que acabou induzindo o sistema ao erro.

Klaus-Robert Müller e Wojciech Samek também mostram que um algoritmo pode fornecer um resultado aparentemente "correto" ao se valer de um percurso lógico equivocado. Os autores oferecem como analogia o caso do cavalo Clever Hans, que ficou célebre no início do século XX por supostamente realizar operações matemáticas básicas, com mais de 90% de acertos. Posteriormente, descobriu-se que, nas apresentações, o cavalo apenas respondia à linguagem corporal do seu treinador, que fazia os cálculos e "soprava" o resultado correto a Hans por meio de sinais. Ou seja, Clever Hans de fato "acertava" os resultados, mas por motivos estranhos à matemática.

De acordo com Müller e Samek (2019, p. 3), o mesmo pode ocorrer com alguns sistemas de IA. Relata-se o caso de um algoritmo classificador de imagens que venceu vários prêmios na área. Mais tarde, foi averiguado que sua predição, muitas vezes, não detectava o objeto principal. Em vez disso, apenas utilizava correlações e dados indiretos para chegar ao resultado. Embora acertasse frequentemente, verificou-se que o modelo reconhecia barcos pela presença de água, trens pela presença de trilhos e até cavalos pela presença de uma marca d'água de direitos autorais embutida na imagem. Esses "preditores Clever Hans" podem ter bom desempenho em cenários de teste, mas certamente falharão quando implementados no mundo real, onde os objetos reconhecidos muitas vezes estão fora do seu contexto original. Müller e Samek (2019, p. 4 – tradução nossa) acrescentam

que, "se o sistema de IA for uma caixa-preta, será muito difícil desmascarar tais preditores".

Portanto, quando um algoritmo produz uma decisão incorreta – ou mesmo correta, mas apoiada em premissas falsas –, estaremos diante da falibilidade, situação em que o sistema de IA não opera da maneira desejável, seja por razões ligadas ao *design* do algoritmo, seja pela forma como os dados são codificados, coletados, selecionados ou usados para treinar o algoritmo. Muitas vezes, a falibilidade tem efeitos inócuos. No entanto, quando uma falha produzida por algoritmos de IA afeta grupos ou indivíduos, potencialmente gerando resultados tendenciosos ou discriminatórios, ela adquire uma dimensão social, razão pela qual passa a ser tratada como *viés algorítmico*, como veremos no próximo tópico.

1.2 Vieses algorítmicos, preconceito e discriminação

Os *vieses algorítmicos* são tendências eventualmente produzidas por um sistema de IA, refletindo, em geral, a predileção humana por determinados valores, devido a fatores sociais e culturais preexistentes à programação e que circundam os projetistas (Rossetti; Angeluci, 2021). A incorporação dessas tendências em um sistema de IA geralmente não se dá deliberadamente pelos programadores, mas pelo treinamento incorreto do algoritmo ou por desdobramentos inesperados do aprendizado de máquina, resultando no comprometimento da neutralidade do sistema.

Considerando que a IA tem sido usada para deliberar a respeito de questões humanas cruciais, a "contaminação" de determinado algoritmo por uma tendência moral poderia reproduzir preconceitos e criar resultados injustos, como privilegiar um grupo de usuários em detrimento de outros (Najibi, 2020). Neste caso, os vieses vão além de uma simples falha, pois podem gerar repercussões sociais graves, como, por exemplo, o reforço de preconceitos sociais relacionados à raça, ao gênero, à sexualidade ou à etnia, gerando discriminações sistemáticas e injustas.

À medida que a IA avança, relatos de preconceito algorítmico se multiplicam a cada ano na literatura científica e seria impossível, neste breve trabalho, esgotar todas as espécies de ocorrências documentadas até o momento. Nos subitens a seguir, selecionamos e resumimos três casos nos quais o *viés algorítmico* pode ter ajudado a produzir decisões sistematicamente discriminatórias. Esses mesmos casos, assim como o exemplo do Husky relatado anteriormente, serão retomados ao final da segunda seção

deste artigo, a partir da perspectiva da XAI, de modo a demonstrar como a inteligência artificial explicável poderia contribuir para a detecção e mitigação dessas falhas e vieses.

1.2.1 Negros indevidamente classificados com altos índices de reincidência

Um exemplo de sistema de IA capaz de produzir resultados discriminatórios é o Compas – *correctional offender management profiling for alternative sanctions* (em português: perfil de gerenciamento corretivo de infratores para sanções alternativas), ferramenta norte-americana utilizada para estimar o risco de reincidência de prisioneiros no país. Uma "classificação de risco" é elaborada pelo sistema com base em um questionário de 137 perguntas feitas ao réu, em seu histórico criminal, e também segundo a base de dados da plataforma. De modo geral, os dados produzidos pelo Compas ajudam a estipular valores de fiança, informam decisões judiciais sobre a liberdade do réu durante o processo, e, em alguns Estados, possuem ainda maior relevância, podendo embasar a sentença criminal (Angwin *et al.*, 2016).

Em 2016, no entanto, uma análise feita pela organização de jornalismo independente *ProPublica* revelou que o algoritmo utilizado no Compas continha vieses discriminatórios. Os autores do estudo analisaram pontuações de mais de 7.000 prisioneiros na Flórida, concluindo que o algoritmo está mais propenso a classificar, equivocadamente, acusados negros como "prováveis reincidentes" e, por outro lado, enquadrar, também de forma equivocada, acusados brancos como "indivíduos com baixo risco de reincidência" (Nunes; Marques, 2018, p. 6).

Os jornalistas da *ProPublica* apuraram ainda que a Northpointe, empresa responsável pelo sistema, não disponibiliza ao público o algoritmo no qual se baseia o índice de reincidência dos detentos, mas tão somente as perguntas feitas ao indivíduo e utilizadas no cálculo, de modo que o réu não sabe por qual motivo possui um alto ou baixo indicador, ou sequer como suas respostas influenciaram na ponderação do resultado final (Angwin *et al.*, 2016).

1.2.2 Preconceito no Google Fotos

Em 2015, um programador negro expôs uma falha no serviço de fotos da Google, que havia rotulado fotos dele e de um amigo negro como "gorilas". À época, a *big tech* veio a público para se declarar "horrorizada" com a

falha, mostrando-se "comprometida" com o fim dos vieses discriminatórios em algoritmos e prometendo "rápida correção" (Simonite, 2018).

Quase três anos depois, em 2018, a revista estadunidense de tecnologia *Wired* testou o Google Fotos usando uma coleção de mais de 40.000 imagens com diversas espécies de animais. Muito embora o algoritmo tenha apresentado desempenho notável ao reconhecer muitas criaturas, como pandas e poodles, o serviço curiosamente relatou "nenhum resultado" para os termos de pesquisa "gorila", "chimpanzé" e "macaco" (Simonite, 2018). Descobriu-se, portanto, que a "solução" dada pela Google ao problema do Google Fotos havia sido simplória: apesar da grande repercussão da denúncia, a empresa apenas removeu gorilas e alguns outros primatas do dicionário do serviço, em vez de refinar os algoritmos de reconhecimento para corrigir suas falhas e vieses.

A exótica solução ilustra as dificuldades que o Google e outras empresas de tecnologia enfrentam no avanço da tecnologia de reconhecimento de imagem, que, todavia, já é aplicado em áreas extremamente sensíveis, como, por exemplo, no controle de migração, no monitoramento de protestos, na vigilância de aeroportos e no combate ao terrorismo.

1.2.3 Discriminação contra mulheres na concessão de crédito

Um recente relatório elaborado por pesquisadores de Oxford mostrou que mulheres empreendedoras tendem a captar menos financiamento junto a acionistas privados, sobretudo devido à resistência dos homens (que são a maioria dos investidores) em financiar empresas lideradas por mulheres. Ainda de acordo com o relatório, muitas empreendedoras entrevistadas revelaram que, a fim de contornar este problema, preferem solicitar dinheiro aos bancos, já que, como estes utilizam algoritmos de pontuação (*score* bancário), seriam automaticamente mais "imparciais" na concessão do crédito (Sako; Parnham, 2021, p. 107).

O problema é que essa suposta neutralidade dos *scores* algorítmicos tem sido colocada em xeque pelos próprios clientes e também por órgãos reguladores norte-americanos. Em 2019, o Departamento de Serviços Financeiros do Estado de Nova York abriu uma investigação sobre denúncias de que o cartão de crédito da Apple estava oferecendo limites de crédito distintos para homens e mulheres. Vários usuários do cartão – incluindo do cofundador da Apple, Steve Wozniak – alegaram, em suas redes sociais, que algoritmos de *score* discriminavam mulheres. O empresário do setor de

tecnologia David Hansson, por exemplo, reclamou que o *Apple Card* deu a ele 20 vezes o limite de crédito que sua esposa obteve, muito embora a renda formal da esposa ultrapassasse a dele. Mais tarde, Wozniak tuitou que a mesma coisa aconteceu a ele e sua esposa, embora não tivessem contas bancárias ou ativos separados. Em sua defesa, o banco Goldman Sachs, que oferece o cartão em parceria com a Apple, alegou que suas decisões de crédito são pautadas essencialmente na qualidade de crédito do cliente, e não em fatores como sexo, raça, idade ou orientação sexual (Natarajan; Nasiripour, 2019).

Apesar de não se ter a certeza do que especificamente ensejou as disparidades relatadas, os problemas com o *AppleCard* revelam um possível enviesamento dos modelos algorítmicos usados em *scores* bancários, que pode estar na origem de efeitos discriminatórios e diferenças de oportunidade entre homens e mulheres.

O que os três exemplos relatados têm em comum? Além do viés algorítmico detectado, que acabou por reproduzir preconceitos de raça e de gênero, todos os sistemas de IA mencionados acima empregavam aprendizado de máquina, cuja sequência preditiva é frequentemente incompreensível – ou seja, opaca – para seres humanos. A *opacidade algorítmica*[6] amplia o desafio de se detectar e corrigir vieses, como veremos a seguir.

1.3 A opacidade pode encobrir falhas e vieses algorítmicos

Na programação tradicional, construir um *software* consistia em redigir um modelo lógico à mão, ou seja, traçar um conjunto de regras que permitiam atingir conclusões a partir do processamento de casos individuais. Tais modelos são, por definição, *interpretáveis*, uma vez que seu código-fonte foi previamente escrito por um desenvolvedor, sendo possível dizer, em cada caso individual, *quais* e *como* as instruções foram acionadas para se chegar a um resultado (*e.g.*: se a renda de um solicitante for inferior a "*y*" por mês, o financiamento será recusado pelo sistema) (Villani, 2019, p. 114).

6 Opacidade algorítmica, de acordo com Harry Surden (2014, p. 158 – tradução nossa), é "qualquer momento que um sistema tecnológico se engaja em comportamentos que, embora apropriados, podem ser difíceis de entender ou prever, do ponto de vista humano".

Por outro lado, algoritmos que incorporam o aprendizado de máquina, como, por exemplo, as *random forests*[7], apenas retornam resultados, sem, contudo, oferecer explicações razoáveis sobre como se chegou a determinada predição. Nesses casos, já que não é possível vislumbrar com clareza o processo decisório por trás do *output*, diz-se que o algoritmo é *opaco* – por constituir uma verdadeira "caixa-preta", incapaz de fornecer explicações razoavelmente compreensíveis para um ser humano. Roberto Confalonieri e seus colegas sintetizam o problema:

> Embora alguns modelos algorítmicos possam ser considerados interpretáveis por design [...] a maioria dos modelos de *machine learning* comporta-se como "caixas-pretas". A partir de uma entrada (*input*), uma "caixa-preta" retornará o resultado [...] sem revelar detalhes suficientes sobre sua lógica interna, resultando em um modelo de decisão *opaco*. (Confalonieri *et al.*, 2020, p. 7 – tradução nossa)

Deve-se pontuar, entretanto, que nem todo aprendizado de máquina conduzirá, necessariamente, à opacidade absoluta. Muitos sistemas de *machine learning* empregam a chamada aprendizagem *supervisionada*, através da qual os programadores "treinam" o algoritmo por meio de exemplos e regras predeterminadas. No caso da aprendizagem supervisionada, a análise deste conjunto de instruções prévias permite conhecer melhor suas etapas de raciocínio e a forma como os dados são analisados. Já no caso da aprendizagem *não supervisionada*, na qual a quantidade de orientações preliminares deixadas pelo desenvolvedor é menor, o algoritmo tem uma operação mais autônoma e, portanto, menos intuitiva para seres humanos. A inteligibilidade do sistema diminui à medida que o algoritmo passa a abrigar não apenas uma, mas múltiplas redes neurais que se sobrepõem, o que comumente ocorre nos algoritmos de *deep learning* (Ghahramani, 2004).

O grande problema é que, em razão da crescente potência e velocidade dos algoritmos de IA – sobretudo os que empregam *deeplearning* –, é quase impossível acompanhar o seu raciocínio, mesmo nos casos em que seu código é aberto. Por exemplo, para reconhecer uma imagem, um classificador pondera milhões de critérios, utilizando milhões de imagens do seu banco de treinamento, as quais, por sua vez, contêm milhões de pixels (4K) (Villani, 2019, p. 114).

---

7 *Random forests* (em português: "florestas aleatórias") são versões mais avançadas das árvores de decisão. Consistem em agrupar grande número de árvores relativamente não correlacionadas, cuja previsão "coletiva" será mais precisa que qualquer uma de suas previsões individuais.

No entanto, em que pese o intricado desafio técnico de se desvendar as "caixas-pretas" algorítmicas, será inevitável enfrentá-lo, tendo em vista o exponencial avanço da IA e seus inúmeros desdobramentos. Se, em alguns casos, as implicações são mínimas, existem, em contrapartida, áreas sensíveis nas quais não se pode admitir a reprodução sistemática desses equívocos, sobretudo se considerarmos que os sistemas de IA possuem a capacidade de tratar problemas em larga escala. Para se ter uma noção dessa escalabilidade, o sistema de IA chamado Victor, em operação no Supremo Tribunal Federal atualmente, é capaz de analisar um processo em 5 segundos, enquanto um servidor, para realizar a mesma tarefa, gasta 44 minutos. Por sua vez, o sistema Athos, utilizado no Superior Tribunal de Justiça, pode analisar até 30 mil processos mensalmente (Andrade, 2021). Nesse sentido, a presença de falhas ou vieses em algoritmos de elevada importância e escalabilidade pode ter implicações de alto risco, como, por exemplo, o enviesamento de centenas ou milhares de análises judiciais.

Desta forma, a incapacidade de se compreender as relações entre dados de entrada (*input*) e dados de saída (*output*) em sistemas de IA pode tornar um sistema de IA uma verdadeira "caixa-preta", à qual não se pode (ou pelo menos não se deveria) entregar decisões cruciais, pela simples impossibilidade de se confiar em um sistema cuja cadeia de raciocínio permanece oculta, podendo encobrir falhas ou vieses discriminatórios. Tome-se o exemplo imaginário de um robô que assassina, de forma aparentemente deliberada, um ser humano. Seria fundamental desvelar a lógica interna que o levou a tal ato – até mesmo para efeitos de responsabilização, se for o caso, dos programadores, do fabricante ou do proprietário daquele robô. Se o algoritmo é uma "caixa-preta", da qual não se pode depreender o processo preditivo interno, seria um grande desafio determinar quando, como e por que o algoritmo errou, informações sem as quais a solução do crime e a atribuição de responsabilidades tornam-se tarefas eminentemente espinhosas.

Não seria o caso, contudo, de afastar os algoritmos de toda e qualquer tomada de decisão, mas sim de abordar a inteligência artificial de forma a incentivar e viabilizar sua explicação, em maior ou menor grau, a depender da sua aplicação final. Essa abordagem deve ter por objetivo a compatibilização da IA com valores sociais, levando-se em conta uma série de questões éticas, entre as quais se destacam a falibilidade, a opacidade, o

viés, a discriminação, a autonomia, a responsabilidade e a privacidade de informações (Rossetti; Angeluci, 2021, p. 7).

Diante de tais reflexões, entendemos que o desenvolvimento de funcionalidades que habilitem a IA a fornecer explicações satisfatórias para seus atos poderia resolver, em parte, o problema da opacidade algorítmica. Não se trata apenas de fomentar a transparência, mas de desenvolver uma postura ativa, que possibilite aos sistemas de IA deixar claras as suas intenções, motivações e o encadeamento causal por trás de uma decisão, notadamente quando esta tem repercussões individuais ou sociais relevantes. Essas propriedades "explicativas" já vêm sendo exploradas no promissor campo da *explainable artificial intelligence (XAI)*, que será tratado a seguir.

2 A INTELIGÊNCIA ARTIFICIAL EXPLICÁVEL

2.1 Explainable artificial intelligence (XAI): conceitos fundamentais

Um sistema inteligente dotado de inteligência artificial explicável (XAI) é aquele que possui *interpretabilidade ou explicabilidade*[8], quer dizer, capacidade para explicar suas predições, por meio de estratégias textuais ou visuais que forneçam compreensão qualitativa sobre seu processo de predição (Ribeiro *et al.*, 2018, p. 2). O sistema de IA explicável está habilitado a fornecer explicações sobre sua operação, tornando seu comportamento mais inteligível para os humanos (Gunning *et al.*, 2019). Significa dizer que um sistema XAI deve estar apto a explicar, de maneira apropriada a um ser humano, a lógica interna de sua predição: o que foi feito, o que está fazendo agora e o que acontecerá a seguir.

A literatura destaca que o nível de detalhes e as características da XAI devem ser estabelecidos levando-se em conta o público-alvo da explicação. Por exemplo, desenvolvedores de *software* podem compreender pequenas redes bayesianas, mas elas são um completo enigma para o usuário leigo (Ribeiro *et al.*, 2016, p. 4). Da mesma maneira, explicações muito básicas devem ser insuficientes para que *experts* revisem ou auditem um algoritmo.

8 Alguns especialistas entendem que *interpretabilidade* e *explicabililidade* são sinônimos. Para outros, contudo, a *interpretabilidade* limita-se à simples transparência do sistema (que sua lógica decisória seja, por si só, ser inteligível para o ser humano), enquanto a *explicabilidade* está ligada a uma postura mais ativa do sistema, à sua capacidade de elaborar e apresentar esclarecimentos humanamente compreensíveis quanto à sua deliberação (Došilović *et al*., 2018, p. 2). Embora essa distinção revele um cuidado conceitual bastante útil para explicações mais técnicas, para o presente trabalho, ela não tem grande relevância, razão pela qual optamos por usar os termos de modo intercambiável.

É primordial, portanto, que o desenvolvimento da XAI não perca de vista o seu consumidor final, devendo fornecer, a depender do destinatário, a) quantidade "apropriada" de informações (nem informações escassas, tampouco em demasia) e b) explicações em linguagem compreensível para o interlocutor.

As primeiras tentativas de gerar XAI remontam às décadas de 1960 e 1970, notadamente por meio dos sistemas Mycin e Centaur, desenvolvidos no âmbito da Universidade Stanford. Contudo, deparou-se com o problema de que as "explicações" engendradas por estes sistemas eram, em verdade, verbalizações das regras, não interpretações consistentes das rotinas ou arquitetura do sistema. Uma expressão formal de "por que fizemos assim" é uma justificativa, não uma explicação (Mueller *et al.*, 2019). Além dessa limitação, há de se considerar o fato de que os sistemas estudados à época (em sua maioria, *sistemas especialistas*[9]) eram substancialmente mais simples do que algoritmos de *machine learning e deeplearning* empregados atualmente. Sendo assim, apesar das tentativas de desenvolvimento da inteligência explicável no passado, ainda persiste o desafio de se fornecer explicações satisfatórias, sobretudo considerando-se a enorme complexificação dos processos algorítmicos trazida pelos avanços da tecnologia computacional, em especial no campo da nanociência.

Como já visto anteriormente, a implementação de sistemas de IA que utilizam aprendizado de máquina tornou-se ponto fulcral de preocupação dos estudiosos, na medida em que alguns modelos algorítmicos constituem verdadeiras "caixas-pretas": de tão complexos, determinados processos preditivos acabam sendo indecifráveis – até mesmo para programadores e técnicos da área. Isso dificulta, por exemplo: a) a experiência e a confiança do usuário final, que pode ser prejudicada por predições equivocadas do sistema; b) o aperfeiçoamento desses modelos por seus desenvolvedores, visto que nem sempre se sabe quando e como a IA falha; c) conformidade legal dessas ferramentas, pois seus detentores estão sujeitos a um risco legal aumentado, decorrente da opacidade e, também, da maior falibilidade de modelos algorítmicos não interpretáveis (Nunes; Andrade, 2021).

9 Os *sistemas especialistas* originais eram *softwares* que objetivavam simular o raciocínio de um humano *expert* em determinada área. Dotados de raciocínio lógico estrito, estes *softwares* logo se mostraram impraticáveis no mundo real, por ignorarem a incerteza. Uma segunda geração de sistemas especialistas passou a empregar técnicas probabilísticas, que, contudo, não salvariam tais *softwares* do fracasso, dado o elevado número de probabilidades existentes e as limitações computacionais da época.

Nessa linha, embora a IA ofereça extraordinárias possibilidades em nosso cotidiano, está bem estabelecido pelos estudiosos do tema que sua opacidade, em alguns casos, não é desejável. De fato, com os algoritmos invadindo nossas vidas diárias, é preciso que reflitam nossas leis e padrões sociais. Diante de "caixas-pretas algorítmicas", o papel da XAI será fundamental para entender, auditar e corrigir esses sistemas, buscando-se permanentemente a sua conformidade ética e legal.

Por fim, vale ressaltar que nem todo sistema de inteligência artificial pressupõe a necessidade de XAI. Molnar (2021) oferece dois exemplos: a) quando o modelo algorítmico e suas predições têm baixo impacto, não havendo desdobramentos sociais, e b) na hipótese em que as aplicações de um determinado sistema já estejam suficientemente estudadas e estabelecidas – como no caso do uso do reconhecimento facial para desbloqueio de celulares.

2.2 A *XAI* na prática

A literatura especializada faz uma distinção entre modelos algorítmicos que são *originalmente interpretáveis* e aqueles que *necessitam ser explicados por meio de técnicas específicas de XAI*. Um modelo "transparente" de aprendizado de máquina é aquele que é explicável por si só, não requerendo técnicas adicionais para que o humano possa compreendê-lo. Em oposição aos modelos "transparentes", existem os "modelos opacos", cuja compreensão irá demandar um processo de explicação adicional, chamado de *explicabilidade post-hoc*. A *explicabilidade post-hoc* se volta para modelos que não são prontamente interpretáveis pelo seu *design*, recorrendo a diversas técnicas de XAI, como explicações de texto, explicações visuais, explicações por meio de exemplos, explicações por simplificação e explicações de relevância de recurso.

Dada a diversidade e profundidade de técnicas XAI, seria impossível discorrer sobre as especificidades de cada uma delas. Mas um exemplo importante nos permite esclarecer o funcionamento da XAI em suas linhas gerais, o que é suficiente para os fins deste estudo. Veja-se o caso da *explicação por relevância de recurso*. Esse método tem por objetivo descrever melhor um algoritmo opaco, enfatizando-se os recursos e variáveis decisivos para o resultado final (*output*) da predição algorítmica. Uma contribuição de destaque são as SHAPs (*SHapley Additive exPlanations*), que propõem uma espécie de "pontuação" para a influência de cada característica preditiva

durante o processamento algorítmico. Por meio das SHAPs, as variáveis usadas no processo preditivo são ordenadas, apresentando-se aquelas que mais influenciam o algoritmo para uma ou outra direção (Lundberg; Lee, 2017).

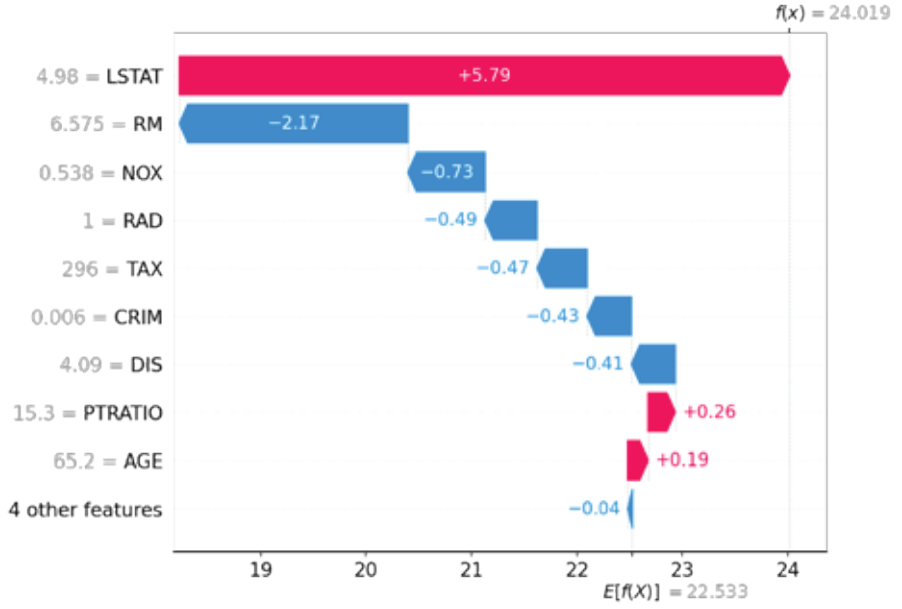


Figura 2: A explicação acima trabalha a ideia de relevância, detalhando como cada recurso contribuiu para o "*output*". Os recursos que "empurram" a previsão para cima são mostrados em vermelho, enquanto os que "empurram" a previsão para baixo são mostrados em azul. Fonte: GitHub (https://github.com/slundberg/shap).

Ao revelar o "peso" dos itens mais decisivos na análise, as SHAPs fornecem informações cruciais sobre o funcionamento do algoritmo, permitindo, por exemplo, que um desenvolvedor identifique uma variável que está sendo sub ou superestimada na predição e, consequentemente, faça regulagens nos pesos de cada variável, de forma a remover vieses e aumentar a precisão dos resultados fornecidos pelo sistema.

3 O VALOR DA EXPLICAÇÃO NO ENFRENTAMENTO DOS VIESES ALGORÍTMICOS

3.1 Detecção de correlações antiéticas

Na primeira seção deste artigo, mostrou-se, por meio dos casos *Compas* e *AppleCard*, como negros e mulheres podem ser prejudicados por

sistemas de pontuação preconceituosos. Na base dessa discriminação, encontram-se algoritmos que, mesmo pretensamente imparciais, podem estar imbuídos da subjetividade de seus criadores ou afetados pela qualidade do treinamento e dos dados fornecidos. No caso de sistemas muito complexos, a opacidade pode agravar e encobrir tais preconceitos, na medida em que dificulta a compreensão da lógica preditiva aplicada. Para resolver tal impasse, a XAI poderia ser implementada, enquanto funcionalidade que ajuda a detectar e corrigir encadeamentos lógicos inerentemente tendenciosos. Afinal, como prelecionam Nunes e Marques (2018, p. 7), a respeito do Compas:

> A ausência de transparência do algoritmo é especialmente crítica nesse caso. Como defender-se de um "índice" sem saber o método de seu cálculo? Como submeter o "índice" ao controle do devido processo constitucional? Por mais que sejam divulgadas as perguntas realizadas, os acusados não sabem como suas respostas influenciam no resultado final (*output*). Dessa forma, a defesa do acusado torna-se impossibilitada por dados matemáticos opacos e algoritmicamente enviesados, mas camuflados, pela "segurança" da matemática, como supostamente imparciais, impessoais e justos.

A XAI pode, inclusive, ajudar a identificar as correlações (não tão óbvias) feitas pelos sistemas de pontuação. Nesse sentido, ainda que o sistema do *AppleCard* não tenha estabelecido uma causalidade direta entre pontuação/raça ou pontuação/gênero, como alega o provedor do serviço, suspeita-se que uma relação indireta pode ter sido traçada a partir de um conjunto de dados mais amplo. Por exemplo, ao longo do casamento, os homens tendem a contrair mais empréstimos em seu nome, em vez de os tomarem em conjunto com suas esposas. Quando não ajustado, esse dado pode levar o algoritmo a inferir que, de modo geral, os homens tomam e honram empréstimos com mais frequência que as mulheres, sendo, portanto, mais confiáveis (Kelion, 2019). De igual forma, o Compas, mesmo sem perguntar a raça do detento, poderia estar absorvendo e considerando essa informação indiretamente, já que, no questionário, existem perguntas que acabam por selecionar indivíduos pobres que são, em sua maioria, negros.

Portanto, por meio de explicações adequadas sobre os processos preditivos dos algoritmos, seria possível identificar (e eventualmente corrigir) o estabelecimento de correlações como essas que, embora indiretas, possuem efeito discriminatório, seja por favorecer homens, como no caso do *AppleCard*, seja por favorecer pessoas de raça branca, no caso do Compas. Constatado, após a emissão da explicação, que um sistema acabou por dar

relevância – mesmo que indiretamente – a um dado racial ou de gênero em seu julgamento, seria o caso, portanto, de recalibrar o modelo algorítmico em questão, evitando-se que ele expresse novos preconceitos em sua pontuação.

3.2 Aperfeiçoando o reconhecimento de imagens

Serengil (2019) demonstra que as SHAPs também são capazes de explicar como um algoritmo distingue entre determinadas emoções por meio do reconhecimento facial. O autor utilizou um conjunto de treinamento de dados chamado FER-2013, que continha imagens de expressões faciais de 28.709 pessoas, sendo capaz de distinguir entre sete categorias (0 = zangado, 1 = nojo, 2 = medo, 3 = feliz, 4 = triste, 5 = surpresa, 6 = neutro).

Na sequência de imagens abaixo, é possível notar como esse método acopla a técnica de *explicação por relevância de recurso* à classificação de imagens, fornecendo informações detalhadas e consistentes sobre o processo de reconhecimento facial de um sistema classificador. De igual forma, a decifração do processo preditivo de um classificador pode ser aplicada em outras áreas, de modo a refinar a precisão do algoritmo (Serengil, 2019).

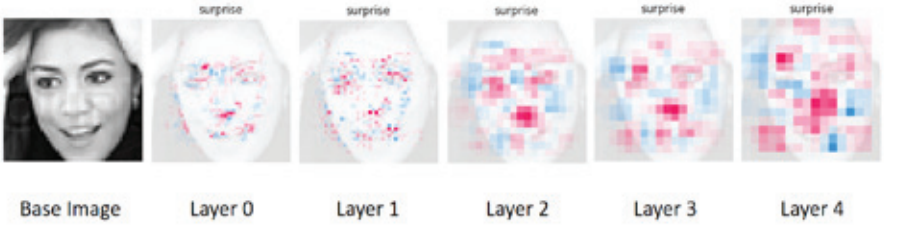


Figura 3: O método SHAP (SHapley Additive exPlanations), utilizado para explicar a detecção algorítmica de uma expressão facial: as primeiras camadas (*layers* 0 e 1) focam nas características do rosto (olhos, nariz, boca etc.), enquanto as camadas seguintes mencionam demais áreas do rosto. Em vermelho, os *pixels* que mais influenciam a previsão, enquanto aqueles com baixa importância estão marcados de azul.

Outro método de explicação que pode ser aplicado a classificadores de imagem é o LIME (*local interpretable model-agnostic explanations*), capaz de explicar como recursos de entrada (*input*) afetam a previsão algorítmica (*output*) (Ribeiro *et al.*, 2016). Resumidamente, esse sistema gera perturbações aleatórias na imagem de entrada, "desligando" e "ligando" alguns *pixels* para encontrar "*superpixels*", ou seja, segmentos de imagem significativos que se assemelham à base de dados catalogada (Arteaga, 2020).

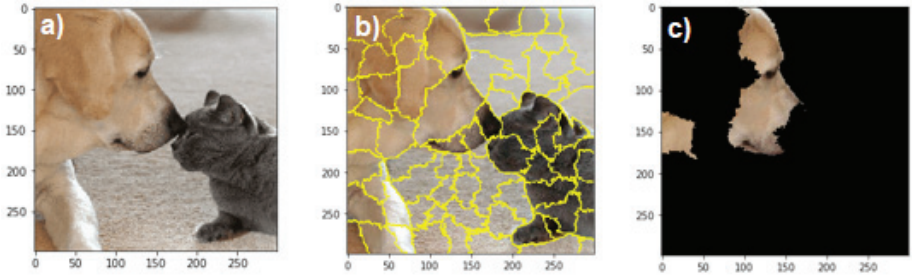


Figura 4: LIME explicando um resultado. Por meio de perturbações na área original (a), "*superpixels*" foram criados (b), e, entre eles, o algoritmo destacou os mais significativos (c), que, comparados com a base de dados, possuem associação mais forte com a raça "Labrador Retriever".

Como se percebe, a explanação acerca do processo preditivo, mesmo que simplificada, pode contribuir para localizar e ajustar falhas. No caso do problema do Husky, por exemplo, o LIME é capaz de revelar, ao usuário da funcionalidade, que a neve adquire significância desproporcional na classificação (Ribeiro *et al.*, 2016). A mesma funcionalidade poderia ser aplicada para ajudar a refinar o reconhecimento do Google Fotos, a fim de identificar e corrigir as grosseiras correlações estabelecidas pelo algoritmo entre *superpixels* presentes na face de negros e de macacos.

Destaca-se ainda que, ao fornecer explicações sobre o processo preditivo, a XAI pode mudar a percepção dos usuários sobre a *confiabilidade* de determinada ferramenta, corrigindo problemas como a "confiança cega" do usuário ou, também, a desconfiança em relação a um algoritmo. Isso ficou demonstrado em estudo desenvolvido por professores da University of Washington que perguntaram a estudantes de programação qual era a sua confiança em um algoritmo classificador de imagem (Ribeiro *et al.*, 2016, p. 9). A princípio, pouco mais de um terço dos indivíduos confiavam no algoritmo. Após receberem a explicação elaborada pelo LIME – revelando que o fundo da imagem continha um peso considerável na classificação –, a confiança no classificador caiu substancialmente (para aproximadamente 10%). Essa evidência chama atenção para a importância de se acoplar explicações aos sistemas, justamente para que usuários não superestimem a confiabilidade de um algoritmo possivelmente impreciso ou enviesado.

CONSIDERAÇÕES FINAIS

Este trabalho mostrou, em sua primeira etapa, como surgem as falhas e os vieses algorítmicos, em especial, analisando-se a ocorrência desses problemas em contextos opacos, ou seja, nos quais se tem pouca ou nenhuma compreensão sobre o processo preditivo empregado por um sistema de IA complexo. Em um algoritmo "caixa-preta", eleva-se o risco de que tais desvios passem despercebidos, podendo ser, inclusive, reproduzidos em larga escala. Se, por um lado, alguns algoritmos têm implicações pouco significativas, por outro, merecem especial atenção aqueles cujos resultados podem ter repercussão individual ou coletiva relevante, a partir de avaliações com base em raça, gênero, sexualidade e etnia capazes de gerar discriminações sistemáticas, como nos casos do Compas e do *AppleCard*.

Diante da constatação de que a opacidade pode encobrir e potencializar falhas e vieses algorítmicos, ponderou-se, na segunda etapa do trabalho, que a inteligência artificial explicável (XAI) poderia auxiliar no combate à opacidade, na medida em que habilita o sistema a fornecer explicações sobre seu próprio processo preditivo. Esse é um processo de transformação da "caixa-preta" algorítmica em uma autêntica "caixa de vidro" – ou seja, transparente, fácil de visualizar e entender – que contribui para a identificação de correlações indesejáveis, estabelecidas no interior do algoritmo, permitindo que desenvolvedores de um sistema rastreiem e corrijam falhas e vieses ali presentes. Ainda, a "caixa de vidro" permite a verificabilidade, auditoria e apuração de responsabilidade quando a IA toma decisões ilegais. Por fim, a XAI também promove a confiança dos usuários e da sociedade na própria inteligência artificial, pois mostra, de maneira geral, quando, como e por que um algoritmo está tomando determinada decisão.

Não se pode esquecer, contudo, que as soluções oferecidas pela inteligência artificial explicável ainda estão situadas no campo experimental, concentrando-se, em sua maioria, em investigações acadêmicas. Nesse sentido, há um percurso considerável até que a XAI seja desenvolvida e implementada satisfatoriamente, de modo a poder beneficiar seu usuário final com explicações úteis e acessíveis. Portanto, será necessário que empresas desenvolvedoras – e, em última análise, seus controladores e investidores – estejam "motivados" a financiar o desenvolvimento de funcionalidades de autoexplicação para sistemas de IA complexos.

Seria ingênuo, no entanto, esperar uma mobilização espontânea por iniciativa das empresas de tecnologia em torno da ética e da transparência

– razão pela qual imaginamos que o verdadeiro catalisador desta motivação repousa, justamente, no debate sobre o direito à explicação. Por isso, a discussão sobre diretrizes relacionadas à explicação e, finalmente, a entrada e estabilização desses direitos nos ordenamentos jurídicos ao redor do globo (a exemplo do que tem ocorrido na União Europeia) podem estabelecer sanções à opacidade algorítmica, incentivando as empresas a tornarem seus *softwares* mais transparentes e interpretáveis.

Apesar de entendermos que a XAI constitui uma via a ser explorada nos próximos anos, é preciso reconhecer que ela não passa, hoje, de uma promessa, que está longe de ser uma unanimidade no mercado. Nada garante seu sucesso, menos ainda de forma ampla e generalizada. Muito provavelmente, será impossível avançar na implementação da explicabilidade em certos domínios mais complexos. Nesses casos, entendemos que expedientes regulatórios mais assertivos serão necessários, especialmente quando estiverem em jogo questões mais sensíveis, como proteção ao meio ambiente ou direitos fundamentais. Esses expedientes podem envolver medidas de controle, de supervisão ou de tutela humana em tempo real, chegando até mesmo ao banimento da tecnologia em determinadas situações.

Esse conjunto de reflexões nos leva a concluir que a compatibilização da inteligência artificial com as leis e valores sociais pressupõe não apenas o acesso, mas também a garantia de compreensão dos processos preditivos empregados em todo sistema de IA ao qual é delegada uma decisão com efeitos significativos – sejam eles a nível privado, como a concessão de um empréstimo pessoal, ou de cunho mais abrangente, como o auxílio à tomada de decisão judicial para sentenciar acusados.

Não se trata somente de incentivar a simples transparência, mas de construir, a nível de políticas públicas, uma postura que reconheça a XAI enquanto requisito para adoção de sistemas de IA em searas mais sensíveis. Considerada a relevância de determinadas decisões, elas devem ser delegadas apenas a algoritmos aptos a esclarecer suas intenções e motivações, explicitando sua análise em linguagem compreensível para o ser humano. Quanto mais importante for uma decisão do ponto de vista social, mais capacitado precisa estar um sistema de IA para fornecer explicações detalhadas, precisas e compreensíveis, para que não pairem dúvidas sobre a sua neutralidade e competência.

REFERÊNCIAS

ALVES, Marco Antônio Sousa. Cidade inteligente e governamentalidade algorítmica: liberdade e controle na era da informação. *Philósophos*, Goiânia, v. 23, n. 2, p. 191-232, 2018. Disponível em: https://www.revistas.ufg.br/philosophos/article/view/52730. Acesso em: 10 ago. 2021.

ANDRADE, Otávio Morato de. Utilizando inteligência artificial para combater a morosidade processual e democratizar o acesso ao judiciário. *Duc in Altum: Cadernos de Direito*. No prelo.

ANGWIN, Julia; LARSON, Jeff; SURYA, Mattu; KIRCHNER, Lauren. Machine Bias. *Pro Publica*, 23 de maio de 2016. Disponível em: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Acesso em: 19 ago. 2021.

ARTEAGA, Cristian. Interpretable machine learning for image classification with LIME: increase confidence in your machine-learning model by understanding its prediction. *Towards Data Science*, 21 de outubro de 2019. Disponível em: https://towardsdatascience.com/interpretable-machine-learning-for-image-classification-with-lime-ea947e82ca13. Acesso em: 26 set. 2021.

BRUNO, Fernanda. *Máquinas de ver, modos de ser*: vigilância, tecnologia e subjetividade. Porto Alegre: Sulina, 2013.

CONFALONIERI, Roberto; COBA, Ludovik; WAGNER, Benedikt; BESOLD, Tarek. A historical perspective of explainable artificial intelligence. *Wires Data Mining and Knowledge Discovery*, v. 11, e1391, 2021. Disponível em: https://wires.onlinelibrary.wiley.com/doi/pdfdirect/10.1002/widm.1391. Acesso em: 19 ago. 2021.

CORTIZ, Diogo. Inteligência artificial: conceitos fundamentais. In: VAINZOF, Rony; GUTIERREZ, Adriei. *Inteligência artificial*: sociedade, economia e Estado. São Paulo: Thomson Reuters, p. 45-60, 2021.

DOŠILOVIĆ, Filip; BRČIĆ, Mario; HLUPIĆ, Nikica. Explainable artificial intelligence: a survey. *41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, p. 210-215, 2018. Disponível em: https://ieeexplore.ieee.org/abstract/document/8400040. Acesso em: 19 ago. 2021.

GHAHRAMANI, Zoubin. *Unsupervised learning*. 16 de setembro de 2004. Disponível em: http://datajobstest.com/data-science-repo/Unsupervised-Learning-Guide-[Zoubin-Ghahramani].pdf. Acesso em: 20 set. 2021.

GUNNING, David; STEFIK, Mark; CHOI, Jaesik; MILLER, Timothy; STUMPF, Simone; YANG, Guang-Zhong. XAI – Explainable artificial intelligence. *Science Robotics*, v. 4, n. 37, 2019. Disponível em: https://robotics.sciencemag.org/content/4/37/eaay7120/tab-article-info. Acesso em: 15 ago. 2021.

FLORIDI, Luciano; TADDEO, Mariarosaria. How AI can be a force for good. *Science*, v. 361 n. 6404, p. 751-752, 2018. Disponível em: https://www.science.org/doi/10.1126/science.aat5991. Acesso em: 26 set. 2021.

KELION, Leo. Apple's "sexist" credit card investigated by US regulator. *BBC News*, 11 de novembro de 2019. Disponível em: https://www.bbc.com/news/business-50365609. Acesso em: 12 ago. 2021.

LUNDBERG, Scott; LEE, Su-In. A unified approach to interpreting model predictions. *arXiv*:1705.07874, 25 de novembro de 2017. Disponível em: https://arxiv.org/abs/1705.07874. Acesso em: 26 set. 2021.

MOLNAR, Christoph. *Interpretable machine learning*: a Guide for Making Black Box Models Explainable. Leanpub, 2021. Disponível em: https://christophm.github.io/interpretable-ml-book/index.html. Acesso em: 13 ago. 2021.

MUELLER, Shane; HOFFMAN, Robert; CLANCEY, William; EMREY, Abigail; KLEIN, Gary. Explanation in Human-AI Systems: a literature meta-review synopsis of key ideas and publications and bibliography for explainable AI. *DARPA XAI Literature Review*, fevereiro de 2019. Disponível em: https://arxiv.org/abs/1902.01876. Acesso em: 14 ago. 2021.

MÜLLER, Klaus-Robert; SAMEK, Wojciech. Towards explainable artificial intelligence. *arXiv*:1909.12072v1, 26 de setembro de 2019. Disponível em: https://arxiv.org/abs/1909.12072. Acesso em: 15 ago. 2021.

NAJIBI, Alex. Racial discrimination in face recognition technology. *Harvard Online: Science Policy and Social Justice*, 24 de outubro de 2020. Disponível em: https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/. Acesso em: 26 set. 2021.

NATARAJAN, Sridhar; NASIRIPOUR, Shahien. Viral Tweet About Apple Card Leads to Goldman Sachs Probe. *Bloomberg*, 9 de novembro de 2019. Disponível em: https://www.bloomberg.com/news/articles/2019-11-09/viral-tweet-about-apple-card-leads-to-probe-into-goldman-sachs. Acesso em: 26 set. 2021.

NUNES, Dierle; ANDRADE, Otávio. A explicabilidade da inteligência artificial e o devido processo tecnológico. *Conjur*, São Paulo, 7 de julho de 2021. Disponível em: https://www.conjur.com.br/2021-jul-07/opiniao-explicabilidade-ia-devido-processo-tecnologico. Acesso em: 26 set. 2021.

NUNES, Dierle; MARQUES, Ana Luiza. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, novembro de 2018. Disponível em: https://www.academia.edu/37764508/INTELIG%C3%8ANCIA_ARTIFICIAL_E_DIREITO_PROCESSUAL_VIESES_ALGOR%C3%8DTMICOS_E_OS_RISCOS_DE_ATRIBUI%C3%87%C3%83O_DE_FUN%C3%87%C3%83O_DECIS%C3%93RIA_%C3%80S_M%C3%81QUINAS_Artificial_intelligence_and_procedural_law_

algorithmic_bias_and_the_risks_of_assignment_of_decision_making_function_to_machines. Acesso em: 26 set. 2021.

RAMOS, Oscar Garcia. "Black box": there's no way to determine how the algorithm came to your decision. *Oscar G. Ramos Blog*, 27 de maio de 2020. Disponível em: https://www.oscargarciaramos.com/blog/9gfzdns1lmwlz58k4w2yxvzjukxp22. Acesso em: 26 set. 2021.

RIBEIRO, Marco Túlio; SINGH Sameer; GUESTRIN, Carlos. "Why should I trust you?": explaining the predictions of any classifier. *arXiv*:1602.04938, 16 de fevereiro de 2016. Disponível em: https://arxiv.org/abs/1602.04938. Acesso em: 26 set. 2021.

ROSSETTI, Regina; ANGELUCI, Alan. Ética algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. *Galáxia*, n. 46, p. 1-18, 2021. Disponível em: http://dx.doi.org/10.1590/1982-2553202150301. Acesso em: 26 set. 2021.

ROUVROY, Antoinette; BERNS, Thomas. Governamentalidade algorítmica e perspectivas de emancipação: o díspar como condição de individuação pela relação? *Revista Eco Pós*, v. 18, n. 2, p. 35-56, 2015. Disponível em: https://revistaecopos.eco.ufrj.br/eco_pos/article/view/2662. Acesso em: 26 set. 2021.

SAKO, Mari; PARNHAM, Richard. *Technology and innovation in legal services*: final report for the solicitors regulation authority. University of Oxford, 2021. Disponível em: https://www.sra.org.uk/globalassets/documents/sra/research/full-report-technology-and-innovation-in-legal-services.pdf?version=4a1bfe. Acesso em: 26 set. 2021.

SERENGIL, Sefik. How SHAP can keep you from black box AI. *Blog Sefik Ilkin Serengil*, 1º de julho de 2019. Disponível em: https://sefiks.com/2019/07/01/how-shap-can-keep-you-from-black-box-ai. Acesso em: 26 set. 2021.

SIMONITE, Tom. When it comes to Gorillas, Google Photos remains blind. *Wired*, 1º de novembro de 2018. Disponível em: https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind. Acesso em: 26 set. 2021.

SUNSTEIN, Cass. *Republic 2.0*. Princeton: Princeton University Press, 2009.

SURDEN, Harry. Machine learning and law. *Washington Law Review*, v. 89, n. 1, 2014. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2417415. Acesso em: 26 set. 2021.

VILLANI, Cédric. *For a meaningful artificial intelligence*: towards a French and European strategy. A parliamentary mission from 8th september 2017 to 8th march 2018. Paris, 2019. Disponível em: https://books.google.com.br/books?id=9cVUDwAAQBAJ&lpg=PP1&hl=pt-BR&pg=PP1#v=onepage&q&f=false. Acesso em: 26 set. 2021.

WELLS, Lindsay; BEDNARZ, Tomasz. Explainable AI and reinforcement learning: a systematic review of current approaches and trends. *Front Artificial Intelligence*, 20 de maio de 2021. Disponível em: https://www.frontiersin.org/articles/10.3389/frai.2021.550030/full. Acesso em: 26 set. 2021.

ZUBOFF, Shoshana. *A era do capitalismo de vigilância*: a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2020.

Sobre os autores:

Marco Antônio Sousa Alves | *E-mail:* marcofilosofia@gmail.com

Professor Adjunto de Teoria e Filosofia do Direito e do Estado da Faculdade de Direito da Universidade Federal de Minas Gerais. Doutor em Filosofia pela UFMG, com estágio de pesquisa na École des Hautes Études en Sciences Sociales (EHESS/Paris). Membro Permanente do Programa de Pós-Graduação em Direito (PPGD/UFMG). Coordenador do Grupo SIGA/UFMG (Sociedade da Informação e Governo Algorítmico).

Otávio Morato de Andrade | *E-mail:* otaviomorato@gmail.com

Mestrando em Direito pela UFMG, com bolsa do CNPq. Especialista em Direito Civil pela PUC-MG. Graduado em Direito (UFMG) e Administração (PUC-MG). Membro-Monitor do Grupo SIGA/UFMG (Sociedade da Informação e Governo Algorítmico).

Data de submissão: 27 de setembro de 2021.
Data de aceite: 10 de janeiro de 2022.

"Falhas de IA" e a Intervenção Humana em Decisões Automatizadas: Parâmetros para a Legitimação pela Humanização

"AI Failures" and Human Intervention in Automated Decision-Making: Parameters for Legitimation through Humanization

MIRIAM WIMMER[1]
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP/DF).

DANILO DONEDA[2]
Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP/SP).

RESUMO: Em um contexto em que diferentes países vêm reconhecendo um direito à intervenção humana face a decisões automatizadas, o artigo tem por objetivo investigar os elementos que podem atrair a necessidade de introdução de parâmetros humanos em processos de decisão impulsionados por sistemas de inteligência artificial. Assim, com base no método hipotético-dedutivo e a partir de pesquisa bibliográfica e documental, o artigo explora diferentes categorias de problemas que podem advir de decisões tomadas por sistemas de IA, concluindo que, em determinados casos, a necessidade de intervenção humana pode ser identificada não apenas com base em critérios de eficiência, mas também pode se constituir em um componente ético em si mesmo. Por outro lado, argumenta-se que determinados parâmetros de sistemas de IA, como o seu nível de transparência e auditabilidade, a explicabilidade das decisões, o seu baixo impacto potencial a direitos e garantias fundamentais e o grau de participação do próprio usuário do sistema na sua configuração e utilização, poderiam mitigar os riscos associados ao "déficit de humanidade" e assim proporcionar que a intervenção humana seja modulada em diferentes níveis de intensidade, mantendo-se o atendimento aos requisitos éticos de decisões legítimas, confiáveis, justas e cognoscíveis, por seres humanos, em seus principais elementos.

PALAVRAS-CHAVE: Inteligência artificial; decisões automatizadas; intervenção humana; ética.

ABSTRACT: In a context in which different countries have come to recognize a right to human intervention on automated decision-making, this paper aims to investigate the elements that may

1 Orcid: https://orcid.org/0000-0001-9210-6651.
2 Orcid: https://orcid.org/0000-0001-9535-3586.

attract the need for human parameters in decision-making supported by artificial intelligence. Thus, based on the hypothetical-deductive method and on bibliographic and documentary research, the article explores different categories of problems that may result from AI decision-making processes, concluding that the need for human intervention can be identified not only based on efficiency criteria, but may also, in certain cases, be considered an ethical component in itself. On the other hand, the paper argues that certain parameters of AI systems, such as their level of transparency, auditability and explainability, their potential low impact for fundamental rights and guarantees as well as the degree of user participation in their configuration and use, could mitigate the risks associated with the "lack of humanity". These parameters could therefore enable the modulation of human participation at different levels of intensity, while meeting the ethical requirements of legitimate, trustworthy and fair decisions, that can be understood by humans in their main elements.

KEYWORDS: Artificial intelligence; automated decisions; human intervention; ethics.

INTRODUÇÃO

Em outubro de 2020, o time escocês de futebol *Scottish Inverness Caledonian Thistle* anunciou, por meio de sua página na internet, que passaria a utilizar, em seu estádio, um novo sistema de câmeras com tecnologia de reconhecimento de imagens baseada em inteligência artificial – IA. Diante das restrições sanitárias que impediam torcedores de comparecer ao estádio, o sistema de filmagem com a tecnologia *ball tracking* permitiria que os torcedores acompanhassem as partidas em tempo real, por meio de uma plataforma de *streaming*, e visualizassem os melhores ângulos das jogadas. Para a frustração dos torcedores, entretanto, a novidade revelou-se um fiasco: as "câmeras inteligentes" reiteradamente confundiram a bola de futebol com a cabeça do árbitro, um senhor completamente calvo (ICTFC Media, 2020).

Tal episódio ilustra, de um lado, a crescente ubiquidade da inteligência artificial, presente em domínios cada vez mais amplos da vida cotidiana; de outro lado, chama atenção para o fato de que tais sistemas frequentemente geram resultados que ensejam preocupações, seja em razão do cometimento de erros objetivos, seja pela existência de fatores que tornem estes resultados questionáveis por apresentar parcialidade, viés, opacidade, tendências discriminatórias, dentre outros aspectos problemáticos. Uma vasta literatura acadêmica explora as consequências jurídicas e as implica-

ções éticas destes fatores no processo decisório de sistemas de inteligência artificial (Pasquale, 2015; O'Neil, 2017; Birhane, 2021).

De fato, ao atribuir a entes automatizados a realização de tarefas que tenham como consequência a participação, total ou parcial, em um processo decisório com impactos relevantes, surge uma demanda clara pela possibilidade de escrutínio dos diversos elementos deste processo decisório. Ainda que sistemas de inteligência artificial possam apresentar enormes vantagens de escala ou mesmo precisão em relação a decisões humanas em diversas searas, sua legitimidade não deriva meramente destas métricas – que podem, inclusive, ser ajustadas mediante critérios utilitaristas ou mesmo instrumentalizadas para determinadas finalidades. Assim, a demanda por uma espécie de instância na qual estas possíveis "falhas" possam ser verificadas e avaliadas acaba, dada a natureza destas decisões, por ser componente ínsito à sua própria natureza e fator imprescindível para a legitimação de seu uso.

Ao mesmo tempo em que se observa a ampliação dos espaços de interação e de decisão mediados por sistemas de IA, é possível verificar, internacionalmente, uma tendência a enfrentar e mitigar os riscos associados a falhas de sistemas algorítmicos. Tal tendência é bastante visível na afluência da pesquisa sobre o tema e no desenvolvimento de parâmetros éticos e normas deontológicas em geral sobre o emprego de sistemas de IA (Baxter, 2021), bem como em diversas estratégias para o uso de IA desenvolvidas por países e organizações, alguns chegando mesmo a flertar com a regulação da matéria[3].

Há, no entanto, uma estratégia observável em normas de proteção de dados pessoais de diferentes países, que é a do estabelecimento de um direito pelo qual indivíduos não estejam sujeitos a decisões baseadas unicamente em processos automatizados, invocando-se, em muitos casos, um direito à intervenção humana, que pode ser materializado, dentre outros mecanismos, por um direito à revisão humana de decisões automatizadas ou por um direito a uma explicação, compreensível por seres humanos, quanto aos seus principais critérios e parâmetros. Apesar das nuances e variações na enunciação desses direitos em diferentes ordenamentos jurídicos, é possível compreender que seu reconhecimento decorre da constatação de

3 No Brasil, no momento da conclusão deste artigo, tramita no Congresso Nacional o Projeto de Lei nº 21/2020, que estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil.

que decisões baseadas total ou predominantemente em sistemas automatizados podem ser consideradas como falhas, necessitando de um tratamento normativo que, antes mesmo de que abordagens específicas sobre IA sejam definidas no patamar regulatório, já estão presentes e, em alguns casos, sedimentadas dentro de modelos regulatórios de proteção de dados.

Diante de tal cenário, o presente artigo tem por objetivo, com base no método hipotético-dedutivo e a partir de pesquisa bibliográfica e documental, investigar os tipos de falhas em que podem incorrer sistemas de IA, com vistas a sugerir critérios que podem atrair, especificamente, a necessidade de introdução de parâmetros humanos em processos de decisão realizados por tais sistemas, assim como parâmetros que podem mitigar os riscos advindos do "déficit de humanidade".

Para os fins deste estudo, adota-se o recorte proposto por Mittelstadt *et al.* (2016) e utilizado também por Tsamados *et al.* (2020), cuja análise é voltada para os algoritmos não apenas como construções matemáticas abstratas, porém por meio de uma perspectiva funcional, sobretudo em vista da forma de sua implementação (artefatos, tecnologias, programas) e configuração (aplicações). Assim, a análise que será empreendida sobre algoritmos e sistemas de IA os considerará enquanto instrumentos usados para converter dados em evidências quanto a um determinado resultado/cenário, que, subsequentemente, é usado para desencadear e motivar uma ação que pode ter consequências que ensejem uma determinada apreciação ética. Será dedicada maior atenção àqueles sistemas que podem tomar ou recomendar decisões, dando apoio à análise humana ou substituindo-a em determinados casos, com base em processos complexos que desafiam a compreensão humana – especialmente os sistemas de inteligência artificial baseados em aprendizado por máquina.

Tem-se como hipótese que, uma vez que as falhas de sistemas de IA podem decorrer não apenas de deficiências formais (desde meros erros de programação ou a utilização de bases de dados de treinamento inadequadas), mas também da impropriedade na tomada de decisões que dependem de percepções, valores ou comportamentos que são, a princípio, incognoscíveis ou impossíveis de serem metrificados e trabalhados por máquinas, a "humanização" de uma decisão pode tornar-se necessária como um componente ético em si mesmo, quando não também sob uma perspectiva de eficiência, baseada nas taxas de erros e acertos de determinado sistema.

Assim, sugere-se que, para além dos parâmetros para atrair a intervenção humana já previstos em normas e declarações internacionais – formulados de maneira ainda muito vaga, como se verá –, devem ser considerados elementos como (i) os riscos e consequências atuais e futuros gerados para os indivíduos e grupos afetados, abrangendo elementos como impactos sobre direitos fundamentais, riscos de discriminação e possibilidade de reversão dos efeitos da decisão, e (ii) a natureza da decisão, em particular no que se refere a decisões em que os juízos de "certo" e "errado" são subjetivos ou em que a decisão deve depender de percepções/valores a princípio incognoscíveis por máquinas. Desta forma, em decisões automatizadas, tanto o seu resultado (as consequências da decisão) quanto seus aspectos procedimentais (a sua natureza) podem ensejar a necessidade do elemento remedial que é a intervenção humana.

Necessário mencionar que, ainda que o direito à intervenção humana (aqui abordado a partir da discussão sobre os direitos à revisão e à explicação de decisões automatizadas) normalmente seja invocado em momento posterior à produção dos efeitos de um sistema automatizado, este pode se materializar por meio da participação de agentes humanos nos processos de tomada de decisão algorítmica de distintas formas e em diferentes momentos do ciclo de vida do produto ou da aplicação em questão. Assim, este artigo não se propõe a identificar, de maneira conclusiva, em qual determinado momento ou de que maneira específica a intervenção humana deva ocorrer para que seja considerada significativa e relevante.

Por outro lado, argumenta-se que determinados parâmetros de sistemas de IA – tais como o grau de transparência, a auditabilidade do sistema, a explicabilidade das decisões tomadas, o seu baixo impacto potencial a direitos e garantias fundamentais, o grau de participação humana exigido ou facultado na sua utilização, entre outros – poderiam mitigar os riscos associados ao "déficit de humanidade". Sugere-se, assim, que a intervenção humana pode ser modulada em diferentes níveis de intensidade, assumindo formas mais brandas, desde que outros mecanismos assegurem o atendimento aos requisitos éticos de decisões legítimas, confiáveis, justas e cognoscíveis em seus elementos, ainda que não necessariamente tomadas por seres humanos.

O artigo está estruturado da forma a seguir descrita. Inicialmente, serão apresentados elementos que permitem compreender as diferentes categorias de problemas associados a decisões por sistemas de IA, abrangendo tanto aspectos relacionados à forma e ao processo de tomada de decisão

por tais sistemas como também aspectos relacionados aos problemas éticos que resultam do uso de tais "decisões" algorítmicas como apoio para decisões humanas ou em substituição a elas. Passa-se, em seguida, a examinar a forma pela qual diferentes jurisdições e organizações internacionais têm estabelecido direitos à intervenção humana em decisões algorítmicas, em particular por meio do direito à revisão de decisões automatizadas e direitos à explicação. Para concluir, com base na discussão apresentada, o artigo sugere parâmetros que podem ser utilizados para identificar a necessidade de intervenção humana, assim como elementos que podem viabilizar a mitigação de tal necessidade ou a modulação de sua intensidade.

1 CATEGORIAS DE PROBLEMAS EM DECISÕES DE IA

As técnicas de decisão algorítmica, cuja relevância, utilidade e ubiquidade no mundo contemporâneo são sempre mais claras, têm sido alvo de escrutínio por diversas razões. Assim, diversos esforços têm sido empreendidos para categorizar os tipos de preocupações éticas associadas a decisões algorítmicas.

Mittelstadt e outros (2016), por exemplo, propõem um mapa conceitual baseado em seis categorias de preocupações éticas ligadas a algoritmos, sendo três de natureza *epistêmica* e duas de natureza *normativa*. No primeiro grupo de preocupações, associado à *qualidade das evidências* produzidas por decisões algorítmicas, encontram-se as seguintes categorias: (i) evidências inconclusivas, que conduzem a problemas de apofenia (*i.e.*, os algoritmos indicam correlações e conduzem à identificação de padrões onde estes não existem verdadeiramente, em fenômeno por vezes também nominado de "correlação espúria"); (ii) evidências inescrutáveis (*i.e.*, opacas ou ininteligíveis); e (iii) evidências mal-orientadas (resultantes da baixa qualidade dos dados de entrada). Quanto às preocupações de natureza *normativa e ética*, os autores indicam os seguintes problemas: (iv) resultados injustos ou discriminatórios e (v) efeitos transformativos, associados à forma como os algoritmos afetam o modo de compreender o mundo, assim como sua organização social e política, com impactos sobre a autonomia humana. Por fim, agravado pelos problemas anteriormente citados, os autores apontam, ainda, para o problema da (vi) dificuldade de *rastreabilidade* das

causas de eventual dano provocado e, consequentemente, de responsabilização dos indivíduos ou organizações envolvidos[4].

Zarsky (2016), por sua vez, em abordagem um pouco mais genérica, identifica duas principais categorias de problemas, ambas relacionadas à natureza opaca e automatizada dos algoritmos, nem sempre passíveis de mitigação por meio de meras medidas de transparência: (i) problemas associados a *ineficiências*, que podem resultar tanto de imprecisões nas bases de dados como também de erros na predição de comportamentos individuais, dada a imprevisibilidade do comportamento humano; e (ii) problemas associados a *injustiças* (transferência injusta de riqueza entre grupos sociais, tratamento discriminatório ou violações à autonomia individual).

Apesar das diferenças nas categorizações acima descritas, é possível visualizar que, em ambas as propostas, há a identificação de duas questões de fundo. A primeira reside na circunstância de que sistemas de IA são intrinsecamente propensos a falhas, particularmente por tomarem decisões com base em métodos e técnicas que não são precisos e nem neutros[5], o que pode levar a resultados inaceitáveis a ponto de serem reconhecidos como falhas (problemas de ineficiência/problemas epistêmicos). Já a segunda questão consiste no fato de que a utilização de "decisões" de IA como apoio para decisões humanas ou em substituição a decisões humanas pode suscitar importantes questionamentos relacionados ao campo da ética, da justiça e da autonomia humana, ensejando igualmente algum tipo de intervenção.

Para os fins da discussão proposta neste artigo – a avaliação de critérios que podem atrair a necessidade de introdução de intervenção humana em processos de decisão automatizados –, entende-se relevante explorar, em maior profundidade, essas duas dimensões.

No que tange à primeira questão, é importante frisar que, ainda que complexos, são já conhecidos e amplamente explorados pela literatura os problemas associados à incorporação de vieses culturais e preconceitos raciais, de gênero e outros em sistemas de aprendizado por máquinas (*machine learning*), que levam a situações em que pessoas integrantes de determi-

4 Para um aprofundamento da discussão com base no mapa conceitual proposto por Mittelstadt *et al.* (2016), v. Tsamados *et al.* (2020).

5 Ao mesmo tempo em que legislações de proteção de dados começavam a se desenvolver, na década de 1970, existia também uma forte crença na objetividade, neutralidade intrínseca e eficiência de sistemas decisionais automatizados, particularmente nos Estados Unidos (Jones, 2017).

nados grupos sociais e étnicos sejam sistematicamente prejudicados por sistemas automatizados de decisão (Pasquale, 2015; O'Neil, 2017). Por serem desenvolvidos com base em dados históricos, que refletem de maneira desigual a diversidade da população ou que incorporam determinados vieses já presentes na sociedade, sistemas desse tipo frequentemente acabam por reproduzir padrões sociais e históricos de injustiça ou discriminação (Edwards; Veale, 2018); assim, as tentativas de agrupar, classificar e predizer o comportamento humano com base nessas técnicas têm se revelado, em alguns casos, bastante problemáticas. Para Birhane (2021), por exemplo, ao tentar impor ordem e identificar padrões no comportamento humano, ferramentas de IA acabam "forçando a determinabilidade, limitando possibilidades e, dessa forma, criando um mundo que se assemelha ao passado", reforçando problemas de discriminação e de injustiça, não raro com consequências particularmente cruéis para grupos marginalizados.

As diferentes abordagens apresentadas para lidar com esses tipos de problemas têm, em muitos casos, focado em ideias como transparência e explicabilidade, de modo a permitir maior visibilidade sobre os critérios que orientam decisões algorítmicas e, assim, viabilizar o seu controle e a sua correção. Tais abordagens partem, em muitos casos, do pressuposto de que é possível aprimorar e aperfeiçoar os sistemas de IA existentes de modo a diminuir as taxas de erros e eliminar eventuais vieses. Com efeito, para determinada corrente de pensamento que, de certa forma, parece carregar o legado das abordagens que propugnavam pela eficácia e objetividade das decisões tomadas por máquina, problemas dessa natureza poderiam ser facilmente corrigidos, caracterizando-se, de maneira explícita, os comportamentos e os resultados aceitáveis e viabilizando-se, assim, que sistemas de aprendizado por máquinas pudessem aprender a desconsiderar fatores discriminatórios de maneira mais efetiva do que humanos. Segundo esse raciocínio, as decisões algorítmicas tenderiam efetivamente a ter maior taxa de acertos do que os próprios seres humanos, particularmente em razão da eliminação de vieses cognitivos de tomadores de decisão humanos (Sunstein, 2018; Kahneman; Sibony; Sunstein, 2021).

De todo modo, a despeito do debate sobre a sua eficácia ou objetividade, é preciso reconhecer que o fato de que as máquinas sejam hoje capazes de realizar tarefas que normalmente são associadas a elevados níveis de discernimento e compreensão humana não significa que os computadores efetivamente possuam discernimento ou compreensão ao realizá-las (Russel; Norvig, 2010, p. 1022) ou, mais ainda, que o façam sob a pers-

pectiva de uma atuação que se possa dizer consciente, conforme veremos. A título de exemplo, muito embora sistemas de IA já sejam hoje capazes, por exemplo, de compor músicas a partir do aprendizado sobre elementos obtidos de obras consagradas, o processo de composição definitivamente é distinto daquele percorrido por um ser humano, que despeja em tal tarefa suas emoções, seu espírito criativo e sua sensibilidade artística – sua humanidade, em síntese.

Assim é que muitas das falhas em que incorrem sistemas de IA – como os sistemas de recomendação que sugerem conteúdos inapropriados, *chatbots* que fornecem respostas absurdas a perguntas formuladas pelo usuário, traduções automatizadas desprovidas de qualquer sentido e sistemas de reconhecimento facial que confundem pessoas com animais – decorrem da ausência de habilidades humanas básicas de percepção do contexto e da cultura no qual estão inseridos[6]. Nessa linha, a história narrada no início deste artigo, sobre um sistema de IA incapaz de diferenciar uma bola de futebol de uma cabeça calva, é um exemplo do chamado Paradoxo de Moravec (Moravec, 1990), segundo o qual habilidades cognitivas que requerem raciocínio lógico, que, no imaginário, costumam ser associadas a um nível elevado de inteligência, são mais facilmente simuladas em um computador do que habilidades simples, como a percepção ou a mobilidade. Desse modo, é mais fácil ensinar um computador a jogar xadrez ou a demonstrar teoremas matemáticos do que ensiná-lo a reconhecer nuances em um tom de voz ou a manipular objetos.

Seguindo essa linha de raciocínio, têm sido levantadas, no campo da filosofia da informação, uma série de objeções à ideia de que sistemas de IA seriam, algum dia, capazes de raciocinar nos mesmos moldes que um ser humano, debate frequentemente apresentado como uma oposição entre a hipótese da IA Fraca, que assevera que as máquinas são capazes apenas de *simular* o pensamento humano, ou seja, agir como se fossem inteligentes, e a hipótese da IA Forte, que afirma que as máquinas efetivamente seriam ca-

6 Para uma interessante descrição de casos históricos de falhas de inteligência artificial, categorizando-as conforme suas consequências, sua intencionalidade, sua evitabilidade e seu estágio de introdução no ciclo de vida do produto, sugere-se a leitura de Scott e Yampolski (2019). Por outro lado, um relato dos sucessos de IA pode ser encontrado em Ganascia (2019), que argumenta que as críticas à IA resultam, em muitos casos, da incompreensão quanto à natureza e aos objetivos da tecnologia, que efetivamente não serve para reproduzir a consciência humana. Para o autor, a IA se traduz em uma disciplina científica que estuda as formas pelas quais a inteligência pode ser decomposta de modo a reproduzir seus diferentes aspectos em computadores. Assim, segundo Ganascia, os falhas de IA resultam, em muitos casos, não de problemas técnicos nos programas de IA, mas residem na sua inadequação social, ou seja, na sua incapacidade de responder às exigências do ambiente social no qual são utilizados.

pazes de pensar e ter autoconsciência, da mesma forma que seres humanos (Russel; Norvig, 2010, p. 1020). As diversas objeções filosóficas apresentadas quanto à ideia de que a existência de sistemas de IA Forte seria possível (Fjelland, 2020) repousam sobre conceitos complexos como consciência, intencionalidade, compreensão e cognição, que não são passíveis de serem explorados em profundidade no contexto deste artigo[7].

Tal discussão, entretanto, permite compreender que uma segunda categoria de problemas associados a decisões de IA, que incluímos no espectro de "falhas", está relacionada não a questões epistêmicas, como deficiências decorrentes de problemas de programação ou a insuficiências nas bases de dados de treinamento, mas sim aos problemas oriundos da delegação a um sistema automatizado da responsabilidade pela tomada de decisões que devem depender de percepções, valores ou comportamentos a princípio não passíveis de conhecimento por máquinas e que consistem basicamente na pletora de emoções e estados de espírito, como bondade, compaixão ou senso de humor, bem como as decisões e tomadas de posição em que os juízos de "certo" e "errado" são subjetivos. Conforme relatam Awad *et al.* (2018), tomando-se como exemplo o conhecido cenário do carro autônomo que, na iminência de um acidente fatal, se vê diante de diferentes possibilidades de decisão – que resultariam na morte do ocupante do carro, de uma criança ou de uma pessoa idosa –, é possível observar que a resposta considerada aceitável pode ter significativas variações entre países, com forte dependência de cultura e de religião – evidenciando a dificuldade de trabalharmos com soluções consideradas mais eficientes, ainda que de ponto de vista utilitarista, que sejam legítimas a despeito da diversidade dos valores culturais envolvidos.

Trata-se de ponto enfatizado, sob outro prisma, também por Birhane (2021), que, ao salientar a "incapacidade de automatizar a ambiguidade", chama atenção para o fato de que seres humanos são sistemas adaptativos complexos, dotados de indeterminabilidade e de uma inerente imprevisi-

7 A título de ilustração, e sem pretensão de esgotar a temática, vale notar que importante campo atual de pesquisa em IA diz respeito à ideia de cognição incorporada, que refuta a ideia de que a cognição poderia se dar por agentes lógicos desconectados de um corpo físico. Segundo essa linha, a cognição é um processo que não funciona de maneira separada de sentidos e estados corporais, que se dão sempre dentro de determinado contexto. Assim, fazendo referência aos trabalhos de Dreyfus (1986) e Clark (1998), Russel & Norvig (2010:1026) apontam que um agente cujo entendimento do termo "cachorro" provém somente de um conjunto limitado de enunciados lógicos (como, por exemplo, "cachorro (x) = mamífero (x)") nitidamente tem uma compreensão mais limitada do animal do que aquela de um agente humano que já observou um cachorro, brincou com o animal e foi lambido por ele.

bilidade, o que contrasta fortemente com a lógica que costuma orientar os sistemas de aprendizado por máquina. Para a autora, as pessoas não podem ser compreendidas fora de seu ambiente (e das normas sociais e das assimetrias de poder ali existentes), de suas trajetórias históricas e de seus valores morais e políticos, que representam elementos cruciais para a sua própria identidade.

Assim é que há domínios que suscitam outro tipo de questionamento: ainda que determinado sistema autônomo atinja um patamar considerado aceitável de taxas de erros e acertos, seria legítimo, eticamente, delegar certos tipos de decisão integralmente a sistemas automatizados, sem intervenção humana relevante?

No campo dos conflitos armados internacionais, por exemplo, sujeitos a regras de direito internacional humanitário concebidas de maneira marcadamente antropocêntrica, debate-se, à luz da chamada "Cláusula de Martens"[8], a própria aceitabilidade de que uma decisão de vida ou morte seja completamente delegada a máquinas, incapazes de mostrar compaixão, de sentir empatia ou de reconhecer a dignidade humana (Wimmer, 2021). Nesse cenário, o que está em discussão, conforme elucida Asaro (2012), não é apenas investigar se um computador, uma máquina ou um processo automatizado são capazes de tomar decisões de vida ou morte e atingir um nível de desempenho considerado aceitável à luz dos preceitos do Direito Internacional Humanitário; mas, sim, se é eticamente aceitável que o ser humano esteja tão distanciado do processo de identificação do alvo a ponto de delegar quase que integralmente a máquinas esse tipo de decisão. Em outras palavras: em determinadas ocasiões, seria legítimo fundamentar decisões em elementos exclusivamente utilitaristas e normativos, extirpando a participação de uma avaliação humana – ainda que esta seja em sentido contraposto?

Debates semelhantes poderiam ser travados com relação a sistemas automatizados operantes em outras searas em que podem existir significativos impactos sobre direitos fundamentais, como na definição de tratamentos médicos ou alocação de leitos em hospitais, no estabelecimento de penas

8 A chamada Cláusula de Martens, também conhecida como o Princípio da Humanidade, encontra-se presente em tratados internacionais e também no Protocolo Adicional às Convenções de Genebra, que estabelece, em seu art. 1º, que as pessoas civis e os combatentes permanecem sob a proteção e o domínio dos princípios do Direito Internacional derivado dos costumes estabelecidos, dos princípios de humanidade e dos ditames da consciência pública.

privativas de liberdade, em decisões críticas de segurança em veículos autônomos e até mesmo, em certas circunstâncias, nas decisões algorítmicas relacionadas à moderação de conteúdos em redes sociais.

É importante ainda registrar que o debate, nestes casos difíceis, não gira apenas em torno da pergunta se, em que circunstâncias e em quais momentos deve ou não existir (ou haver a possibilidade de requerer) a intervenção humana; trata-se, também, de definir que tipo de intervenção humana seria qualitativamente apta a suprir o "déficit de humanidade" em tomadas de decisões sobre as quais recai uma elevada carga moral ou que produzem um significativo impacto sobre direitos fundamentais.

Embora diferentes estratégias jurídicas e regulatórias venham sendo debatidas para lidar com tais desafios, é possível identificar, conforme mencionado anteriormente, uma tendência, em instrumentos internacionais e em legislações de proteção de dados pessoais de variados países, de previsão de direitos associados à intervenção humana em determinados tipos de decisões automatizadas. Diferentes modalidades de intervenção humana previstas em outras jurisdições e os critérios usados para desencadear o exercício de tais direitos serão, a seguir, brevemente examinados.

2 MODALIDADES DE INTERVENÇÃO HUMANA EM DIFERENTES JURISDIÇÕES

Pelo fato de que as propostas de regulamentação abrangente da inteligência artificial são ainda embrionárias, é no campo da proteção de dados pessoais que se pode observar, mais claramente, tentativas de promover a introdução de elementos "humanos" em decisões tomadas automaticamente. A ideia de que, pelo menos em alguns casos, indivíduos não devem estar sujeitos a decisões tomadas unicamente com base em algoritmos tem sido endereçada por meio de diferentes estratégias jurídicas, com destaque para o reconhecimento de direitos (i) à explicação sobre as características, critérios e consequências de decisões automatizadas e (ii) à revisão de tais decisões.

Os direitos à explicação e à revisão são, aqui, tomados em um vetor genérico. Há, por exemplo, tanto leituras como proposições do primeiro que ora se assemelham mais a um direito à explicação, ora mais a um direito à informação. E, ainda, há uma interessante ressonância entre ambos: muito embora o direito à explicação não pressuponha necessariamente a participação humana, trata-se de ferramenta que permite que um ser humano compreenda e exerça controle sobre os principais aspectos relacionados

a decisões automatizadas e pode também representar elemento essencial para o próprio exercício do direito de revisão. Da mesma forma, o direito à revisão, tomado em toda a sua tessitura, implica o reconhecimento de um inafastável componente informativo ligado ao próprio direito à explicação, visto que uma revisão somente se legitima quando é capaz de explicitar os critérios e vetores que a inspiraram – o que nada mais é que um elemento de "explicação" sobre a sua *ratio*.

Verifica-se, ademais, o surgimento de crescente consenso acerca do tema em organismos internacionais. Observe-se, por exemplo, que, muito embora as Diretrizes da OCDE sobre privacidade (2013) não tenham abordado o tema da participação humana em decisões automatizadas, essa ideia é claramente enunciada na Recomendação do Conselho da OCDE sobre Inteligência Artificial (2019), que estabelece a necessidade de que organizações e indivíduos envolvidos com o desenvolvimento e a utilização de sistemas de IA implementem mecanismos e salvaguardas, como a capacidade para a determinação humana, que sejam apropriados ao contexto e consistentes com o estado da arte, de modo a assegurar o respeito à lei, aos direitos humanos e aos valores democráticos[9]. Ao abordar o tema da transparência e da explicabilidade, a Recomendação indica ainda a necessidade de que sejam fornecidas informações relevantes, apropriadas ao contexto e consistentes com o estado da arte, que permitam que aqueles afetados por um sistema de IA possam compreender o resultado e contestá-lo, com base em informações simples e facilmente compreensíveis sobre os fatores e a lógica que serviram de base para a predição, recomendação ou decisão.

Também a versão modernizada da Convenção para a Proteção das Pessoas relativamente ao Tratamento Automatizado de Dados de Caráter Pessoal, conhecida como Convenção 108+, explicitamente prevê direitos relacionados à participação humana em decisões automatizadas[10].

Mais recentemente, em novembro de 2021, a Conferência Geral da Unesco aprovou a Recomendação sobre a Ética da Inteligência Artificial, na qual se enuncia que, em cenários em que as decisões de sistemas de IA pos-

9 Os valores e direitos citados são a liberdade, a dignidade, a autonomia, a privacidade e a proteção de dados, a não discriminação, a igualdade, a diversidade, a equidade, a justiça social e os direitos trabalhistas.

10 "*Article 9 – Rights of the data subject. 1. Every individual shall have a right: a. not to be subject to a decision significantly affecting him or her based solely on an automated processing of data without having his or her views taken into consideration; [...] 2. Paragraph 1.a shall not apply if the decision is authorised by a law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights, freedoms and legitimate interests*."

sam produzir impactos irreversíveis ou de difícil reversão, ou que envolvam decisões de vida ou morte, deve existir determinação humana final. Ademais, a Recomendação indica que a supervisão humana compreende não apenas a supervisão humana individual, mas também a supervisão pública, conforme apropriado. A Recomendação afirma, por fim, que um sistema de IA nunca pode substituir a responsabilidade e *accountability* humanos em última instância (Unesco, 2021).

Não há dúvidas, entretanto, de que os direitos que decorrem da afirmação da necessidade de intervenção humana carecem de maior densidade conceitual, de modo que existe ainda ampla margem para disputas interpretativas sobre seu conteúdo e sobre a sua concreta forma de fruição em diferentes jurisdições (Jones, 2017). Merece destaque, em particular, a controvérsia acerca da própria viabilidade de exercício de um direito à explicação face à complexidade e opacidade dos modelos usados por sistemas de IA para chegarem a determinadas decisões.

De modo a exemplificar duas abordagens relevantes para a presente discussão, passa-se a examinar como o tema tem sido abordado na Europa e no Brasil.

2.1 Europa

O Regulamento Europeu de Proteção de Dados Pessoais – RGPD[11] estabelece, em seu art. 22, que:

> O titular dos dados tem o direito de não ficar sujeito a nenhuma decisão tomada exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar.

Observa-se, portanto, que o RGPD tem como ponto de partida uma vedação à tomada de decisões exclusivamente automatizadas com efeitos jurídicos ou similarmente significativos. As exceções a tal regra são definidas também pelo art. 22, admitindo-se as decisões exclusivamente automatizadas nos seguintes casos: (i) quando o tratamento for necessário para a execução ou celebração de um contrato; (ii) quando o tratamento for autorizado pelo direito da União ou do Estado-Membro, sendo necessário,

11 Regulamento (UE) nº 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva nº 95/46/CE (Regulamento Geral sobre a Proteção de Dados).

entretanto, que a legislação preveja medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular de dado; ou (iii) quando o tratamento for baseado no consentimento explícito do titular de dados.

Ademais, ainda que o titular de dados consinta com o tratamento automatizado ou que este seja necessário para a celebração de um contrato, conforme hipóteses (i) e (iii) *supra*, o Regulamento determina a necessidade de estabelecimento, pelo controlador, de "medidas adequadas para salvaguardar os direitos e liberdades e legítimos interesses do titular dos dados, designadamente o direito de, pelo menos, *obter intervenção humana por parte do responsável*, manifestar o seu ponto de vista e contestar a decisão".

O art. 22 do RGPD tem sido objeto de intensa discussão, notadamente no que diz respeito à existência e à extensão de um possível "direito à explicação" que dele decorreria. Conforme relatam Souza, Perrone e Magrani (2021), a controvérsia resulta do fato de que, muito embora o Considerando 71 do RGPD mencione o direito do titular de dados de obter uma explicação sobre a decisão automatizada e contestá-la[12], tal previsão acabou por não ser explicitada no corpo do art. 22 do Regulamento. Ao mesmo tempo, o art. 15 do RGPD estabelece que, no caso de decisões automatizadas, o titular tem o direito de obter informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento. Tais circunstâncias conduziram a uma ampla gama de interpretações quanto ao escopo do dispositivo, como a posição de Wachter, Mittelstadt e Floridi (2017) no sentido de que o GDPR não estabelece um amplo direito à explicação, mas um limitado direito de ser informado quanto às funcionalidades de um sistema de decisão automatizado; e o posicionamento no sentido contrário, pela concretude do direito à explicação no ordenamento europeu, esposada por Selbst e Powles (2017), que derivam este direito diretamente da previsão do RGPD quanto ao direito a uma informação significativa (*meaningful information*) acerca da lógica utilizada na decisão e suas consequências.

Edwards e Veale (2018) descrevem outras polêmicas que têm cercado a interpretação do art. 22 do RGPD. O que seria uma "decisão" para fins

12 Confira-se: "Em qualquer dos casos, tal tratamento deverá ser acompanhado das garantias adequadas, que deverão incluir a informação específica ao titular dos dados e o direito de obter a intervenção humana, de manifestar o seu ponto de vista, *de obter uma explicação* sobre a decisão tomada na sequência dessa avaliação e de contestar a decisão".

do RGPD? Esta incluiria, por exemplo, o envio de publicidade dirigida, que poderia ser facilmente ignorada pelo titular? No que consistiriam os citados "efeitos na esfera jurídica" ou outros efeitos que afetem significativamente o titular? E, por fim: o que seria uma decisão tomada "exclusivamente" com base no tratamento automatizado, e que nível de participação humana afastaria a incidência da norma?

Ao analisar o art. 22 do RGPD para fins de estabelecer orientações sobre decisões automatizadas e a definição de perfis, o Grupo de Trabalho do Artigo 29 para a Proteção de Dados (WP29, 2018) acabou por fixar uma interpretação ampla quanto ao seu escopo, indicando que decisões tomadas "exclusivamente com base" no tratamento automatizado são aquelas em que não há uma supervisão humana relevante. Assim, segundo a interpretação do grupo, o controlador não pode se eximir da incidência do artigo fabricando uma intervenção humana meramente simbólica; esta deve ser realizada "por alguém com autoridade e competência para alterar a decisão e que, no âmbito da análise, deverá tomar em consideração todos os dados pertinentes"[13].

Ao tempo em que reconheceu que o art. 22 abrange apenas as situações em que há impactos graves para o titular de dados, o Grupo de Trabalho do Artigo 29 também interpretou, de maneira abrangente, a expressão "que produza efeitos na sua esfera jurídica ou que o afete significativamente de forma similar". Para o colegiado, quando o RGPD faz referência a "efeitos na esfera jurídica", é preciso que a decisão em questão afete concretamente direitos de alguém[14]. Por outro lado, segundo o grupo de trabalho, quando a norma introduz a expressão "ou que o afete significativamente de forma similar", há uma abertura do âmbito de incidência da norma, que passa a ser aplicável quando os efeitos de uma decisão automatizada sejam suficientemente grandes ou importantes para merecerem atenção – vale dizer, quando a decisão puder: (i) afetar, de maneira significativa, as circunstâncias, o comportamento ou as escolhas das pessoas em causa;

13 Na mesma linha, a autoridade de proteção de dados britânica indica que a regra em questão se refere a situações em que não há influência humana sobre o processo decisório. Assim, um processo pode ser considerado totalmente automatizado se um ser humano apenas inserir os dados a serem processados, sendo a decisão tomada por uma máquina. Um processo não será considerado totalmente automatizado se um ser humano avaliar e interpretar os resultados de uma decisão automatizada antes de aplicá-la a um indivíduo (ICO, 2018).

14 Por meio, por exemplo, da limitação à liberdade de associação, ao direito de voto ou à possibilidade de mover ações judiciais; ou quando forem afetados o estatuto jurídico de uma pessoa ou os seus direitos no âmbito de um contrato.

(ii) provocar um impacto prolongado ou permanente no titular dos dados; ou (iii) ensejar a exclusão ou discriminação das pessoas.

O art. 22.2 (b) do RGPD indica a possibilidade de que os Estados-Membros definam, por lei, hipóteses autorizativas de decisões exclusivamente automatizadas, sendo necessário, entretanto, estabelecer medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses dos titulares de dados. Como nota Malgieri (2019), é possível visualizar abordagens bastante heterogêneas entre os países europeus na implementação da norma; para os fins deste artigo, interessa chamar atenção para duas das quatro abordagens descritas pelo autor. A primeira, caracterizada como "procedimental", foi adotada pelo Reino Unido, Irlanda e Eslovênia, e inclui a definição de procedimentos a serem adotados por controladores de dados quando houver tomadas de decisão automatizadas, como a obrigação de notificação ao titular sobre a decisão automatizada e o estabelecimento de procedimentos para exercício de um direito de revisão. A segunda abordagem, chamada de "proativa", inclui o estabelecimento de salvaguardas novas e mais detalhadas, como o direito de conhecer os métodos e critérios usados em sistemas específicos ou o direito de receber informações específicas sobre a implementação do sistema de decisão algorítmica – segundo Malgieri, é esse o caso da França e da Hungria[15].

Por fim, vale notar que, especificamente no campo da prevenção, investigação, detecção ou repressão de infrações penais ou execução de sanções penais, a Diretiva (UE) nº 2016/680 estabelece, em seu art. 11, a proibição de decisões tomadas exclusivamente com base no tratamento automatizado, incluindo a definição de perfis, que produzam efeitos adversos na esfera jurídica do titular dos dados ou que o afetem de forma significativa. Tal regra é excepcionada unicamente em casos que sejam autorizados pelo direito da União ou do Estado-Membro ao qual o controlador está sujeito, desde que a legislação preveja garantias adequadas dos direitos e

15 No caso da França, é importante também chamar atenção para a decisão de 2018 do Conselho Constitucional acerca da constitucionalidade da lei de proteção de dados francesa, que aborda o tema da transparência algorítmica e da explicação de decisões automatizadas. V. Conseil Constitutionnel, Décision nº 2018-765 DC du 12 juin 2018, §71: "*le responsable du traitement doit s'assurer de la maîtrise du traitement algorithmique et de ses évolutions afin de pouvoir expliquer, en détail et sous une forme intelligible, à la personne concernée la manière dont le traitement a été mis en œuvre à son égard. Il en résulte que ne peuvent être utilis*és, comme fondement exclusif d'une décision administrative individuelle, des algorithmes susceptibles de réviser eux-mêmes les règles qu'ils appliquent, sans le contrôle et la validation du responsable du traitement".

liberdades do titular dos dados, pelo menos o direito de obter a intervenção humana do responsável pelo tratamento.

2.2 Brasil

Também no Brasil, a Lei Geral de Proteção de Dados Pessoais – LGPD[16] introduziu previsões semelhantes, embora formuladas de maneira ainda mais aberta do que se verificou na Europa e com ênfase menos direta na participação humana. Com efeito, o art. 20 da LGPD estabelece que "[o] titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses [...]". O § 1º do mesmo artigo, por sua vez, determina que o controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial[17].

Observa-se, de imediato, que, apesar das supramencionadas críticas quanto à sua vagueza, o RGPD europeu apresenta parâmetros mais detalhados do que aqueles trazidos pela LGPD no que tange ao exercício dos direitos associados à intervenção humana em decisões algorítmicas. Para além das perguntas já enunciadas com relação ao RGPD – o que é uma decisão e no que consiste uma decisão tomada unicamente com base em tratamento automatizado –, a LGPD enseja ainda outra: a que tipo de "interesses" afetados a lei estaria a se referir no art. 20, que justificariam a incidência da norma? Quaisquer interesses afetados por uma decisão automatizada, por mais triviais que fossem, seriam aptos a ensejar um direito à revisão?[18]

É interessante notar que, enquanto a discussão na União Europeia claramente tem sido pautada pela defesa da participação humana nos processos de decisão capazes de produzir efeitos jurídicos ou de afetar significativamente os interesses do titular de dados, boa parte do debate no Brasil

16 Lei nº 13.709, de 14 de agosto de 2018.

17 Merece destaque a posição de Renato Leite Monteiro (2018) no sentido de que a LGPD teria expandido o conceito de um direito à explicação em comparação com o RGPD, trazendo um rol de proteções mais amplo que aquelas previstas na regulação europeia. Tal interpretação decorre, sobretudo, do entendimento do autor de que, diferentemente do que ocorre no regulamento europeu, a LGPD teria previsto a possibilidade de que, caso o processo automatizado tenha por finalidade formar perfis comportamentais ou se valha de um perfil comportamental para tomar uma decisão subsequente, haveria também a possibilidade de o titular ter acesso aos dados anonimizados utilizados para enriquecer tais perfis. Outra diferença significativa, segundo o autor, residiria no fato de que, diferentemente da LGPD, o RGPD veio a limitar o direito de oposição do titular nos casos em que a base legal para o tratamento é a execução de um contrato ou o consentimento.

18 Levantando questionamentos similares, v. Mulholland e Frajhof (2019) e Frazão (2018).

ainda gira em torno do questionamento se o citado direito à revisão acarretaria sempre e necessariamente uma revisão por um ser humano, ou se uma revisão automatizada de uma decisão automatizada seria então admissível perante a lei. A controvérsia decorre das diversas alterações ao dispositivo durante a sua tramitação pelo Congresso Nacional, que resultaram na eliminação da expressão "por pessoa natural", que integrava a versão original do art. 20[19].

Conforme delimitado anteriormente, o objetivo deste estudo não é de propor uma interpretação definitiva dos referidos artigos, que são trazidos a lume apenas para exemplificar a complexidade do debate e evidenciar um ponto a ser abordado na próxima seção: a carência de critérios claros que possam ser considerados para atrair a necessidade de intervenção humana no caso de decisões tomadas por sistemas automatizados, em especial no caso de utilização de sistemas de inteligência artificial.

3 PARÂMETROS PARA ATRAIR OU ATENUAR A NECESSIDADE DE INTERVENÇÃO HUMANA

Da discussão precedente, é possível compreender que o principal critério eleito tanto pelo RGPD como pela LGPD para viabilizar a contestação de decisões automatizadas diz respeito, essencialmente, aos seus *efeitos*. O RGPD estabelece, como critério para a intervenção humana face a decisões automatizadas, a produção de *efeitos na esfera jurídica* do titular ou outros que o afetem *significativamente* de forma similar; já a LGPD, ao tratar da revisão de decisões automatizadas (ainda que sem explicitar a participação humana) e do direito do titular de obter informações acerca dos critérios e procedimentos utilizados, menciona, de maneira ainda mais genérica, as decisões "que *afetem seus interesses*", inclusive aquelas destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade[20]. Conforme visto anteriormente, a imprecisão da lingua-

19 A redação original da LGPD foi alterada pela Medida Provisória nº 869, de 2018, suprimindo-se a referência à pessoa natural. Na tramitação da medida provisória pelo Congresso, foi novamente incluída menção à revisão por pessoa natural, por meio da inserção de um parágrafo ao dispositivo. Esse parágrafo foi, por fim, objeto de veto presidencial, baseado no seguinte argumento: "A propositura legislativa, ao dispor que toda e qualquer decisão baseada unicamente no tratamento automatizado seja suscetível de revisão humana, contraria o interesse público, tendo em vista que tal exigência inviabilizará os modelos atuais de planos de negócios de muitas empresas, notadamente das *startups*, bem como impacta na análise de risco de crédito e de novos modelos de negócios de instituições financeiras, gerando efeito negativo na oferta de crédito aos consumidores, tanto no que diz respeito à qualidade das garantias, ao volume de crédito contratado e à composição de preços, com reflexos, ainda, nos índices de inflação e na condução da política monetária".

20 Interessante notar que, no projeto de lei de conversão da MP 869, chegou a ser aprovada a seguinte redação pelo Congresso Nacional, posteriormente vetada pelo Presidente da República, para definir a necessidade de revisão humana: "Art. 20. [...] § 3º A revisão de que trata o caput deste artigo deverá ser realizada por

gem jurídica impõe o estabelecimento de critérios interpretativos adicionais que permitam avaliar se, quando e em que circunstâncias a intervenção humana seria necessária para legitimar a decisão adotada.

Nesse sentido, convém recordar que existe um amplo espectro de aplicações de IA e uma diversidade quase infindável de contextos em que tais tecnologias podem ser utilizadas, desde tradutores de textos usados em ambientes profissionais a armamentos letais usados em conflitos armados. Assim, embora seja possível visualizar, nas extremidades de tal espectro, circunstâncias em que seja mais evidente a necessidade de maior ou menor participação humana no processo decisório, há um sem-número de situações intermediárias que podem suscitar fundadas dúvidas.

Outro aspecto a se considerar é que também a intervenção humana pode assumir diferentes matizes. Ainda que se afaste a validade de uma intervenção humana meramente simbólica, conforme preconizado pelo Grupo de Trabalho do Artigo 29 para a Proteção de Dados (WP29, 2018), permanecem diversas questões relacionadas à definição das características para que uma participação humana seja significativa e adequada a cada contexto.

Assim, cabe, em primeiro lugar, considerar que a intensidade da intervenção humana pode variar bastante em função do grau de delegação de decisões à máquina. Para ilustrar o raciocínio, tome-se como exemplo uma aeronave completamente autônoma, em que há apenas indicação da origem e do destino do voo, sendo todos os demais parâmetros definidos de maneira automatizada, sem qualquer intervenção humana. Agora, imagine-se, como segundo exemplo, que essa mesma aeronave seja capaz de fazer a viagem autonomamente, mas permanece sob supervisão de um piloto humano ao longo de todo o trajeto. Por fim, considere-se, como terceiro cenário, a situação de uma aeronave em que o sistema autônomo é responsável apenas pelas etapas de pouso e decolagem, ficando todas as demais definições e ações a cargo do piloto. Embora em todos os casos descritos o voo tenha sido viabilizado por um sistema automatizado, é evidente que a

pessoa natural, conforme previsto em regulamentação da autoridade nacional, que levará em consideração *a natureza* e o *porte da entidade ou o volume de operações de tratamento de dados*". Como se pode observar, a redação vetada previa, como critérios para a revisão humana, elementos essencialmente empresariais e econômicos, afastando-se, assim, da discussão que correlaciona o direito à revisão a um mecanismo de proteção da autonomia individual.

intensidade da participação humana foi radicalmente distinta em cada um dos cenários.

Ainda quanto às formas de intervenção humana, um segundo ponto a ser considerado diz respeito à efetiva capacidade humana (intelectual, emocional, motora) de (re)avaliação substantiva de decisões sugeridas por sistemas de IA. Em muitos casos, a própria opacidade do processo decisório de sistemas de IA dificulta a identificação de erros que possam ter sido cometidos. Em outras circunstâncias, o volume de dados tratados por meio de decisões automatizadas é tão elevado que não há condições para que uma avaliação caso a caso seja sequer realizável por um humano. Por fim, é preciso ainda considerar a amplamente difundida visão de que as decisões automatizadas são sempre mais precisas e mais corretas do que as decisões humanas, o que acaba gerando, para o decisor humano, um ônus argumentativo por vezes excessivo ou mesmo intransponível para não adotar a solução sugerida pelo sistema automatizado.

Um terceiro ponto a ser considerado diz respeito ao momento da intervenção humana. Vale destacar que a questão não se limita a saber se determinada intervenção humana deva ocorrer antes ou após a produção de efeitos pela decisão automatizada, mas requer, inclusive, uma avaliação quanto ao grau de afastamento temporal entre a decisão humana e a decisão de IA que faz tal decisão humana ainda ser relevante. No caso de utilização de armas autônomas letais, por exemplo, pode-se compreender que quanto maior o intervalo de tempo entre o momento em que uma arma autônoma é ativada por um operador humano e o momento em que a arma seleciona e ataca um alvo, maior o risco de que as premissas que embasaram a decisão humana não mais sejam válidas, especialmente quando seu uso se dá em ambientes dinâmicos ou densamente povoados (Lawand, 2020).

À luz do exposto, pode-se compreender que os critérios comumente previstos em normas e em documentos internacionais para atrair direitos associados à intervenção humana, anteriormente apresentados, são ainda muito vagos e pouco sistematizados para que se possa, em casos concretos, definir, de maneira precisa, em quais circunstâncias e de que maneira a intervenção humana pode ser exigível.

Decerto, a dificuldade de estabelecimento de critérios rígidos para tal avaliação decorre também do fato de que a análise deve necessariamente levar em consideração o contexto da decisão automatizada. Ainda assim, entende-se que é possível, com base nas reflexões anteriores, apontar ele-

mentos que podem trazer maior densidade para tais critérios, a partir da consideração dos diferentes tipos de "falhas" que podem ser cometidas em decisões automatizadas com apoio em sistemas de IA – passíveis, como se viu, de cometerem erros não apenas em razão de deficiências de programação ou falhas nas bases de dados de treinamento, mas também em razão de sua incapacidade de tomada de decisões em que os juízos de "certo" e "errado" são subjetivos ou em que a decisão deve depender de percepções, valores ou comportamentos a princípio não passíveis de conhecimento por máquinas.

Assim, sugere-se que a avaliação quanto à necessidade, forma e momento de intervenção humana, baseada na avaliação dos efeitos da decisão automatizada, deve incluir, ao menos, um juízo quanto: (i) aos riscos e consequências atuais e futuros gerados para os indivíduos e grupos afetados, abrangendo elementos como impactos sobre direitos fundamentais, riscos de discriminação e possibilidade de reversão dos efeitos da decisão; e (ii) à natureza da decisão, em particular no que se refere a decisões em que os juízos de "certo" e "errado" são subjetivos ou em que a decisão deve depender de percepções/valores a princípio incognoscíveis por máquinas. Desta forma, em decisões automatizadas, tanto o seu resultado (as consequências da decisão) quanto seus aspectos procedimentais (a sua natureza) podem ensejar a necessidade do elemento remedial que é a intervenção humana.

É válido apontar que esse tipo de avaliação pode ser concretizada pela realização de avaliações de impacto que, conforme elucida Mantelero (2018), devem ser pautadas pelos direitos e valores em jogo e, consequentemente, podem ter abordagens específicas para o contexto (por exemplo, avaliações de risco no campo da saúde podem levar em consideração elementos distintos daquelas empreendidas no campo da segurança pública). Para o autor, a adoção de uma abordagem orientada a valores impõe, adicionalmente, um foco sobre o impacto social do uso de dados, abrangendo potenciais resultados negativos para direitos e princípios fundamentais e levando em conta, também, as consequências éticas e sociais do tratamento de dados[21]. Nessa linha, Mantelero argumenta que modelos de avaliação de

21 É importante observar, entretanto, que existem determinados tipos de riscos que podem ser identificados em avaliações desse tipo que não serão necessariamente solucionados por meio da intervenção humana – mencione-se, a título exemplificativo, riscos associados à coleta excessiva de dados pessoais ou riscos associados à formação de perfis por meio de inferências.

impacto já existentes ou em discussão – como os já conhecidos relatórios de impacto à proteção de dados, os relatórios de impacto a direitos humanos, ou, ainda, os relatórios de impactos éticos – poderiam evoluir para um modelo mais completo, abrangendo a avaliação de impactos quanto a direitos humanos, ética e sociedade (*Human Rights, Ethical and Social Impact Assessment – HRESIA*), incluindo, em alguns casos, a previsão de consultas a comitês independentes de especialistas para apoiar avaliações éticas mais complexas.

A avaliação dos riscos e dos possíveis efeitos, atuais e futuros, de uma decisão automatizada movida por um sistema de IA pode também ser útil para que se permita concluir em qual momento uma participação humana relevante é necessária e de que maneira ela deve ocorrer, caso ela de fato seja imprescindível – ou seja, se, em determinado caso, um direito à intervenção humana *a posteriori* seria mecanismo adequado e suficiente para viabilizar a proteção dos direitos dos indivíduos afetados; se haveria necessidade de participação humana *a priori*, ou seja, antes que a decisão automatizada produzisse efeitos concretos; ou, ainda, se a supervisão humana significativa seria necessária ao longo de todo o ciclo de vida do sistema (incluindo etapas como pesquisa, desenvolvimento, utilização, manutenção, operação, monitoramento, avaliação etc.). Nesse sentido, a irreversibilidade dos efeitos da decisão automatizada certamente é elemento central a ser considerado, pois, embora um direito à explicação pudesse eventualmente apoiar demandas de reparação por danos experimentados, pouco sentido haveria em prever o direito à revisão de uma decisão cujos efeitos são irreversíveis.

Por outro lado, é possível, sem pretensão de exaustividade, vislumbrar a existência de outros parâmetros de sistemas de IA que, ao conferirem maior legitimidade às decisões tomadas, têm o condão de suprir, em certa medida, o "déficit de humanidade" e assim proporcionar que sejam consideradas formas mais brandas de participação humana, com a condição de que sejam atendidos plenamente os requisitos éticos de decisões legítimas, confiáveis, justas e cognoscíveis, por seres humanos, em seus principais elementos.

Nesse sentido, tomando-se como ponto de partida a constatação de que a transparência é, efetivamente, um princípio fundamental da proteção de dados pessoais tanto no Brasil como em outros países, um dos aspectos centrais a ser considerado diz respeito justamente ao grau de transparência/

opacidade do sistema em questão[22] e à compreensibilidade do processo decisional. Os casos de algoritmos inteiramente determinísticos, em que os resultados são sempre previsíveis e passíveis de compreensão, suscitam questões éticas e jurídicas de muito mais fácil resolução do que aquelas geradas por sistemas do tipo "caixa preta forte" (Bathaee, 2018), caracterizados por processos de tomada de decisão em que não há forma (a) de se determinar como a IA chegou a uma decisão ou previsão, (b) de saber qual informação foi determinante para o resultado alcançado pela IA, ou (c) de obter um *ranking* das variáveis processadas pela IA em sua ordem de importância.

Nesse sentido, tomando o requisito da transparência pelo seu viés da "explicabilidade" e abordando a questão dos modelos que, por sua complexidade, geram decisões que não são intrinsecamente interpretáveis, Maranhão, Cozman e Almada (2021) esclarecem que técnicas de Explainable AI, ou XAI, têm sido utilizadas para viabilizar a geração de explicações quanto à forma pela qual decisões foram tomadas, permitindo, assim, que tais decisões sejam simuladas, contrastadas com outras alternativas plausíveis e, eventualmente, questionadas pelos indivíduos afetados quanto a possíveis efeitos antijurídicos, por meios judiciais ou extrajudiciais.

Um segundo parâmetro a ser considerado diz respeito ao impacto da decisão automatizada sobre direitos fundamentais. Neste sentido, pode-se trabalhar com uma matriz de risco na qual as decisões automatizadas podem ser classificadas em torno dos seus impactos potenciais aos direitos fundamentais, pela qual aquelas com menor potencial de impacto podem, em conjunto com outros fatores, justificar modelos mais genéricos de intervenção humana ou mesmo permitir que outros elementos, tal como a transparência, uma vez verificados a contento, possam mesmo eximir a necessidade de que esta intervenção seja direta. A partir desta métrica, o aumento potencial dos impactos aos direitos fundamentais presentes em decisões automatizadas pode ensejar a necessidade de que formas mais específicas de intervenção humana sejam observadas, compreendendo, inclusive, hipóteses nas quais o recurso às decisões automatizadas seja desaconselhado de todo.

22 Nesse sentido, veja-se a proposta de classificação de Burrell (2016), que identifica três diferentes formas de opacidade algorítmica: (i) a opacidade intencional, como mecanismo corporativo ou institucional de autoproteção e ocultação; (ii) a opacidade decorrente do fato de que escrever e ler código computacional é uma habilidade limitada a especialistas; e (iii) a opacidade que resulta do descasamento entre os procedimentos matemáticos de algoritmos capazes de aprendizado e os estilos humanos de interpretação semântica.

Por fim, um terceiro aspecto a ser considerado para modular a necessidade ou intensidade de intervenção humana nos processos de tomada de decisão automatizados diz respeito às possibilidades de participação do próprio indivíduo afetado pela decisão na configuração e nos resultados do sistema – que, a depender de sua efetividade, poderiam mesmo descaracterizar o conceito de decisão tomada unicamente com base no tratamento automatizado de dados pessoais. De fato, a ideia de *design* centrado no usuário tem ganhado força nos debates éticos sobre sistemas de IA, particularmente no contexto de sistemas de recomendação, levando à sugestão de que alguns problemas éticos decorrentes de exposição a conteúdo inapropriado, por exemplo, podem ser endereçados, inclusive, por meio do estabelecimento de filtros especificados pelo próprio usuário (Milano *et al.*, 2020). Assim, as possibilidades e o efetivo grau de participação do indivíduo afetado pela decisão automatizada na sua configuração – por exemplo, permitindo que ele afaste ou calibre a relevância de determinados resultados ou critérios de decisão – podem ajudar a orientar uma decisão quanto à forma e ao momento em que o direito à intervenção humana (por meio de um direito à explicação ou à revisão, por exemplo) pode ser exercido, ou mesmo afastar a sua exigibilidade[23].

CONCLUSÃO

Considerações sobre a necessidade e a modulação de intervenções humanas em procedimentos que incluem decisões automatizadas são, de certa forma, variações de um dos eixos clássicos dos estudos – e de divagações! – sobre inteligência artificial, que é a própria natureza da atividade realizada pelos sistemas e artefatos que incluem essa técnica: eles realmente "pensam" como os humanos? Desta atividade pode decorrer uma legitimidade que permita considerar tais decisões em paralelo com aquelas realizadas por humanos?

A efetiva utilização e implementação de sistemas decisionais baseados em inteligência artificial trouxe, necessariamente, uma boa dose de pragmatismo para este debate. Ainda que diversos de seus elementos ontologicamente mais relevantes continuem demandando intensa reflexão,

23 Por outro lado, é importante não perder de vista que as soluções centralizadas no usuário possuem, também, determinadas limitações, que são expostas em detalhes no estudo de Milano *et al.* (2020). Para os autores, um ponto importante a ser considerado diz respeito ao fato de que, embora tal abordagem possa estimular a transparência, ela tem também o efeito colocar integralmente sobre os ombros do usuário a responsabilidade pela proteção de direitos e pela utilidade da aplicação.

surgiu a necessidade de respostas imediatas para problemas e dilemas que se colocam crescentemente diante de nós a partir deste crescente recurso a estes sistemas. A partir destas demandas mais concretas e imediatas, outras abordagens se impõem. Por exemplo, saber se sistemas de IA podem ou não "pensar" como os humanos acaba sendo, nestas dinâmicas, uma discussão secundária – algo equivalente a questionar se submarinos sabem nadar (Russel; Norvig, 2010, p. 1021), por exemplo, diante da constatação de que tais sistemas, na prática, já fazem, efetivamente, escolhas e deliberam acerca de questões que impactam, de maneira profunda, a vida das pessoas.

Sendo, portanto, fato que o impacto destes sistemas decisionais automatizados já se percebe em inúmeras circunstâncias, também é necessário que as ferramentas aplicadas para equalizar seus efeitos levem em conta os óbices em se contar exclusivamente com ações específicas dos cidadãos, titulares de dados, para legitimá-las ou não. A depender de como sejam empregados, estes mecanismos – seja a revisão, explicação ou outras formas de intervenção humana – acabam colocando um fardo pesado sobre os ombros do titular de dados[24].

A esse argumento se somam outros, como o de que eventualmente a revisão ou a explicação *a posteriori* de decisões automatizadas podem vir tarde demais, a depender dos efeitos já produzidos, como nos casos, por vezes limítrofes porém concretos, de danos referentes à liberdade ou à integridade física por conta destas decisões.

Assim, a discussão sobre intervenção humana deve, necessariamente, contemplar outras questões relacionadas a estruturas de governança e supervisão de decisões automatizadas de forma mais ampla e matizada conforme as circunstâncias específicas e os riscos específicos aos direitos e garantias das pessoas envolvidas.

Conclui-se, assim, que os mencionados direitos à revisão ou à explicação têm fundamento, sobretudo, na introdução de um componente humano no processo decisório automatizado, em particular, no caso de sis-

24 "*The right to an explanation is only one tool for scrutinizing, challenging, and restraining algorithmic decision making. While it has rhetorical strength in demanding transparency to enable user challenge, it has serious practical and conceptual flaws [...] In short, a legal right to an explanation may be a good place to start, but it is by no means the end of the story. Rights become dangerous things if they are unreasonably hard to exercise or ineffective in results, because they give the illusion that something has been done while in fact things are no better. It is instructive here to compare the history of consent to sharing of data, which has moved in the online world from a real bulwark of privacy to something most often described as meaningless or illusory*." (Edwards; Veale, 2018)

temas de IA, com o objetivo de viabilizar que a decisão seja submetida a um controle de legitimidade que envolve critérios e parâmetros compreensíveis e contestáveis por seres humanos. Constatando-se que os direitos em questão, portanto, são essencialmente instrumentos que servem para endereçar os problemas da confiança na máquina, pode-se compreender que outros parâmetros de legitimação são também importantes e podem relativizar a forma e a intensidade da necessidade de participação humana em processos decisórios automatizados.

Justamente a conjugação de diferentes instrumentos de "humanização" permite compreender que a revisão humana não é elemento imprescindível em todos os casos, por ser instrumental para o elemento ético de ter decisões legítimas, confiáveis, justas e cognoscíveis em seus elementos, ainda que não necessariamente revistas por humanos.

REFERÊNCIAS

ASARO, P. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. *International Review of the Red Cross*, 94 (886), p. 687-709.

AWAD, E.; DSOUZA, S.; KIM, R.; SCHULZ, J.; HENRICH, J.; SHARIFF, A.; BONNEFON, J; RAHWAN, I. The moral machine experiment. *Nature*, v. 563, p. 59-64, 2018.

BATHAEE, Yavar. The artificial intelligence black box and the failure of intent and causation. *Harvard Journal of Law & Technology*, Vol. 31, N. 2, p. 890-938, Spring 2018.

BAXTER, Kathy. Ethical AI frameworks, tool kits, principles, and certifications – Oh my! 2021. Disponível em: https://blog.einstein.ai/frameworks-tool-kits-principles-and-oaths-oh-my/. Acesso em: dez. 2021.

BIRHANE, A. The impossibility of automating ambiguity. *Artif Life (2021)*, 27 (1): 44-61. Disponível em: https://doi.org/10.1162/artl_a_00336. Acesso em: ago. 2021.

BURREL, J. How the machine "thinks": understanding opacity in machine learning algorithms. *Big Data & Society*, January-June 2016. Disponível em: journals.sagepub.com/doi/pdf/10.1177/2053951715622512. Acesso em: 26 dez. 2018.

DONEDA, D.; MENDES, L.; DE SOUZA, C.; ANDRADE, N. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. *Pensar*, v. 23, n. 4, 2018. Disponível em: https://periodicos.unifor.br/rpen/article/view/8257. Acesso em: dez. 2021.

EDWARDS, L.; VEALE, M. Enslaving the algorithm: from a "right to an explanation" to a "right to better decisions"? *IEEE Security & Privacy*, 16(3), 46-54, 2018, doi:10.1109/MSP.2018.2701152.

FJELLAND, R. Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, 7, 10 (2020), https://doi.org/10.1057/s41599-020-0494-4. Disponível em: https://www.nature.com/articles/s41599-020-0494-4. Acesso em: dez. 2021.

FRAZÃO, A. Controvérsias sobre direito à explicação e à oposição diante de decisões automatizadas. *Revista Jota*, 12 dez. 2018. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/controversias-sobre-direito-a-explicacao-e-a-oposicao-diante-de-decisoes-automatizadas-12122018. Acesso em: nov. 2021.

GANASCIA, J. G. Epistemology of AI revisited in the light of the Philosophy of Information. *Knowledge, Technology & Policy*, v. 23, p. 57-73, 2010.

ICO – INFORMATION COMMISSIONER'S OFFICE. *Detailed Guidance on Automated Decision-Making and Profiling* (2018). Disponível em: https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling-1-1.pdf. Acesso em: nov. 2021.

ICTFC MEDIA. ICTTV Live Streaming From Caledonian Stadium. *Inverness Caledonian Thistle Football Club*, 16 out. 2020. Disponível em: https://ictfc.com/icttv-live-streaming-from-caledonian-stadium. Acesso em: set. 2021.

JONES, M. L. Right to a Human in the loop: political constructions of computer automation & personhood from data banks to algorithms. *Soc. Stud. of Sci.*, v. 47, 2017.

KAHNEMAN, D.; SIBONY, O.; SUNSTEIN, C. R. *Noise*. A flaw in human judgement. London: William Collins, 2021.

LAWAND, K. International law, incluing IHL, on LAWS: is there a need for a new protocol? In: *Rio Seminar on Autonomous Weapons Systems*, Rio de Janeiro, Naval War College. Brasília: Funag, 2020.

MALGIERI, G. Automated decision-making in the EU Member States: the right to explanation and other "suitable safeguards" in the national legislations. *Computer Law & Security Review*, 35, 2019. Disponível em: https://www.researchgate.net/publication/334359463_Automated_decision-making_in_the_EU_Member_States_The_right_to_explanation_and_other_suitable_safeguards_in_the_national_legislations. Acesso em: set. 2021.

MANTELERO, A. AI and Big Data: a blueprint for a human rights, social and ethical impact assessment. *Computer Law & Security Review*, Vol. 34, Issue 4, p. 754-772, August 2018.

MARANHÃO, J.; COZMAN, F. G.; ALMADA, M. Concepções de explicação e do direito à explicação de decisões automatizadas. In: VAINZOF, R.; GUTIERREZ, A. (Org.). *Inteligência artificial*: sociedade, economia e Estado. São Paulo: Thomson Reuters, 2021.

MILANO, S., TADDEO, M.; FLORIDI, L. Recommender systems and their ethical challenges. *AI and Society*, 2020, 35, 957-967. Disponível em: https://link.springer.com/article/10.1007/s00146-020-00950-y. Acesso em: set. 2021.

MITTELSTADT, B. D.; ALLO, P.; TADDEO, M.; WACHTER, S.; FLORIDI, L. The Ethics of Algorithms: Mapping the Debate. *Big Data & Society 3*, 2016, https://doi.org/10.1177/2053951716679679.

MONTEIRO, Renato Leite. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? *Artigo Estratégico 24*. Rio de Janeiro: Instituto Igarapé, dez. 2018. p. 9-10. Disponível em: https://igarape.org.br/existe-um-direito-a-explicacao-na-lei-geral-de-protecao-de-dados-no-brasil/. Acesso em: set. 2021.

MORAVEC, H. *Mind children*. The future of robot and human intelligence. Cambridge/London: Harvard University Press, 1990.

MULHOLLAND, C.; FRAJHOF, I. Z. Inteligência artificial e a Lei Geral de Proteção de Dados Pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de machine learning. In: FRAZÃO, A.; MULHOLLAND, C. (Org.). *Inteligência artificial e o Direito*: ética, regulação e responsabilidade. São Paulo: Thomson Reuters, p. 272-290, 2019.

O'NEIL, Cathy. *Weapons of math destruction*: how Big Data increases inequality and threatens democracy. New York: Random House Audio, 2017.

OECD – ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT. *The OECD Privacy Framework*. Paris, 2013. Disponível em: https://www.oecd.org/digital/ieconomy/privacy-guidelines.htm. Acesso em: set. 2021.

_____. *Recommendation of the council on artificial intelligence*. Paris, 2019. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: set. 2021.

PASQUALE, Frank. *The black box society*: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015.

RUSSEL, Stuart J.; NORVIG, Peter. *Artificial intelligence*. A modern approach. New Jersey: Pearson Education Inc., 2010.

SCOTT, P. J.; YAMPOLSKI, R. V. Classification Schemas for artificial intelligence failures. *Delphi – Interdisciplinary Review of Emerging Technologies*, Vol. 2, Issue 4, Pages 186 – 199, 2019. Disponível em: https://delphi.lexxion.eu/article/DELPHI/2019/4/8. Acesso em: ago. 2021.

SELBST, A. D.; POWLES, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, vol. 7, n. 4, p. 233-242. Disponível em: https://ssrn.com/abstract=3039125. Acesso em: dez. 2021.

SOUZA, C. A.; PERRONE, C.; MAGRANI, E. O direito à explicação entre a experiência europeia e a sua positivação na LGPD. In: BIONI, B. R.; DONEDA, D.; SARLET, I. W.; MENDES, L. S.; RODRIGUES JR, O. L. (Org.). *Tratado de Proteção de Dados Pessoais*. São Paulo: Forense, p. 243-270, 2021.

SUNSTEIN, Cass. Algorithms, correcting biases (December 12, 2018). No prelo. *Social Research*. Disponível em: https://ssrn.com/abstract=3300171. Acesso em: set. 2021.

TSAMADOS, A.; AGGARWAL, N.; COWLS, J.; MORLEY, J.; ROBERTS, H.; TADDEO, M.; FLORIDI, L. *The ethics of algorithms*: key problems and solutions (28 de julho de 2020). Disponível em: https://ssrn.com/abstract=3662302. Acesso em: set. 2021.

UNESCO – UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION. *Recommendation on the Ethics of Artificial Intelligence*. Adotada em 24 de novembro de 2021 pela Conferência Geral da Unesco. Disponível em: https://en.unesco.org/artificial-intelligence/ethics#recommendation. Acesso em: nov. 2021.

WACHTER, S.; MITTELSTADT, B.; FLORIDI, L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. (December 28, 2016). *International Data Privacy Law* (2017). Disponível em: https://ssrn.com/abstract=2903469. Acesso em: set. 2021.

WIMMER, M. Inteligência artificial e conflitos armados internacionais: o problema das armas autônomas letais. In: VAINZOF, R.; GUTIERREZ, A. (Org.). *Inteligência artificial* – Sociedade, economia e Estado. 1. ed. São Paulo: Revista dos Tribunais, 2021.

WP29 – Grupo de Trabalho do Artigo 29 para a Proteção de Dados. *Orientações sobre as decisões individuais automatizadas e a definição de perfis para efeitos do Regulamento (UE) 2016/679*. Adotadas em 3 de outubro de 2017, com a última redação revista e adotada em 6 de fevereiro de 2018. Disponível em: https://ec.europa.eu/info/law/law-topic/data-protection_pt. Acesso em: set. 2021.

ZARSKY, T. Z. (2016) The trouble with algorithmic decisions: an analytic road map to examine efficiency and fairness in automated and opaque decision making. *Science, Technology & Human Values*, 41(1), p. 118-132, 2016. Disponível em: https://law.haifa.ac.il/images/documents/0162243915605575.pdf. Acesso em: set. 2021.

Sobre a autora e sobre o autor:

Miriam Wimmer | *E-mail:* miriam.wimmer@yahoo.com.br

Doutora em Políticas de Comunicação e Cultura pela Faculdade de Comunicação da Universidade de Brasília (UnB). Mestre em Direito Público e Graduada em Direito pela Universidade do Estado do Rio de Janeiro – UERJ. Ex-Bolsista do Programa Internacional da Universidade de Waseda, com Distinção Acadêmica. Certificada como Especialista em Proteção de Dados Pessoais (Europa) pela International Association of Privacy Professionals (CIPP/E). Professora do Corpo Permanente do Mestrado Profissional em Direito do IDP-Brasília e Professora Convidada em diversas instituições de ensino de nível superior. Servidora Pública desde 2007, integrante da carreira de Especialista em Regulação de Serviços Públicos de Telecomunicações da Anatel. Ocupou diferentes cargos de direção no Ministério das Comunicações – MC e no Ministério de Ciência, Tecnologia, Inovações e Comunicações – MCTIC, onde coordenou a elaboração da Estratégia Brasileira para a Transformação Digital e atuou na propositura da Estratégia Brasileira de IA. É, atualmente, Diretora da Autoridade Nacional de Proteção de Dados – ANPD.

Danilo Doneda | *E-mail:* danilo.doneda@idp.edu.br

Doutor em Direito Civil pela Universidade do Estado do Rio de Janeiro (UERJ). Professor no Instituto Brasiliense de Direito Público (IDP). Membro indicado pela Câmara dos Deputados para o Conselho Nacional de Proteção de Dados e Privacidade. Diretor do CEDIS/IDP (Centro de Estudos de Internet e Sociedade). Membro do Conselho Diretor da IAPP (International Association of Privacy Professionals). Foi Coordenador-Geral na Secretaria Nacional do Consumidor do Ministério da Justiça, onde coordenou a redação do Anteprojeto de Lei de Proteção de Dados, a base da Lei Geral de Proteção de Dados. Membro da Comissão de Juristas da Câmara dos Deputados para redação de Projeto de Lei sobre Proteção de Dados nos setores de segurança pública e investigação criminal. Membro do Grupo de Trabalho sobre proteção de dados e informações judiciais do Conselho Nacional de Justiça. Membro dos Conselhos Consultivos do Projeto Global Pulse (ONU) e do Projeto Criança e Consumo (Instituto Alana). Foi Pesquisador Visitante na Autoridade Garante para a Proteção de Dados em Roma (Roma, Itália), na Università degli Studi di Camerino (Camerino, Itália), e no Instituto Max Planck para Direito Privado Comparado e Internacional (Hamburgo, Alemanha). Parte do seu trabalho está disponível em: www.doneda.net.

Data de submissão: 30 de setembro de 2021.
Data de aceite: 15 de dezembro de 2021.

Microtrabalho e Inteligência Artificial: Desafios à Fruição da Dignidade Humana em Meio à Aprendizagem de Máquina

Microwork and Artificial Intelligence: Challenges to the Realization of Human Dignity in the Context of Machine Learning

DENISE PIRES FINCATO[1]
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

JULISE CAROLINA LEMONJE[2]
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS).

RESUMO: O artigo trata do tema das plataformas de microtrabalho que ofertam serviços para o desenvolvimento e aperfeiçoamento da inteligência artificial. Busca-se, a partir das condições de trabalho e características descritas pela literatura científica, articular o cenário do microtrabalho à concretização do princípio e valor dignidade humana nesse contexto. Assim, por meio de pesquisa bibliográfica, do método de abordagem hipotético-dedutivo e da interpretação sociológica, tecem-se considerações sobre os impactos do microtrabalho na fruição da dignidade humana. Para isso, o artigo inicia com a apresentação de noções conceituais acerca do trabalho em plataformas digitais para, em seguida, apontar as principais características do microtrabalho e sua relação com o desenvolvimento de inteligência artificial. Após, articula-se o impacto das condições mapeadas na concretização da dignidade humana. Conclui-se que a tendência a ocultar a existência de microtarefas desempenhadas por humanos na produção de inteligência artificial, assim como a gestão dessa atividade humana a partir de uma lógica de máquina, enseja a instrumentalização dos trabalhadores e restrições à autonomia do microtrabalhador, que se vê dissociado do valor social do seu trabalho e desconhece elementos essenciais da atividade desempenhada.

PALAVRAS-CHAVE: *Crowdwork*; dignidade da pessoa humana; inteligência artificial; microtrabalho; plataformização do trabalho.

ABSTRACT: The article deals with the subject of microwork platforms that offer services for the development and improvement of artificial intelligence. From the working conditions and characteristics

1 Orcid: http://orcid.org/0000-0002-1339-9343.
2 Orcid: http://orcid.org/0000-0001-5742-4590.

described by the scientific literature, the aim is to relate the microwork context to the realization of the principle and value of dignity of the human person in this scenario. Thus, through bibliographical research and the method hypothetical-deductive of approach and sociological of interpretation, considerations are made about the impacts of microwork on the exercise of human dignity. For this, the article begins with the presentation of conceptual notions about digital labor platform and then points out the main characteristics of microwork and its relationship with the development of artificial intelligence. Afterwards, the impact of the characteristics mapped on the realization of human dignity is associated. It is concluded that the tendency to hide the existence of microtasks performed by humans in the production of artificial intelligence, as well as the management of this human activity from a machine logic, causes the instrumentalization of workers and restrictions on the autonomy of the microworkers, who see themselves dissociated from the social value of their work and are unaware of essential elements of the activity performed.

KEYWORDS: Crowdwork; dignity of the human person; artificial intelligence; microwork; labor platform.

INTRODUÇÃO

Os impactos do desenvolvimento da inteligência artificial nas relações de trabalho geram incertezas acerca da existência futura de demanda por mão de obra humana, culminando em entraves vinculados à ocupabilidade e previdência social. As repercussões da *machine learning* na seara laboral, todavia, extrapolam a questão da falta de procura por trabalho humano e entrelaçam-se, também, ao surgimento de novas ocupações humanas vinculadas ao aperfeiçoamento de inteligência artificial.

Para além da demanda por empregados com altos e específicos conhecimentos na área da tecnologia, a indústria da inteligência artificial é geradora de procura por trabalhadores que se sujeitem ao desempenho de atividades pouco qualificadas e cuja remuneração é marcada pela incerteza. Nesse sentido, as plataformas de microtrabalho oferecem às empresas que desenvolvem aprendizagem de máquina, mormente a inteligência baseada em estatísticas, a realização de microtarefas por uma multidão de trabalhadores dispersos, visando sanar as necessidades de geração, anotação e verificação de dados.

Observa-se, entretanto, que as atividades vinculadas às plataformas de microtrabalho se dão à margem da regulação juslaboral clássica, tendo

suas condições determinadas de maneira unilateral pelas plataformas. Diante desse cenário, emerge a preocupação com a efetiva fruição da dignidade humana em uma conjuntura predominantemente dominada pela máquina, e em que elementos humanos tendem a ser silenciados e invisibilizados.

Tendo em vista a ainda pouco conhecida conjuntura do microtrabalho e os desafios que acompanham este contexto, o presente artigo tem como questão norteadora a identificação dos impactos das particularidades do trabalhado desempenhado junto a plataformas digitais na fruição, pelo trabalhador, da dignidade que lhe é intrínseca. Objetiva-se, portanto, tecer considerações e mapear as adversidades que repercutem na concretização do princípio e valor constitucional máximo, em atenção à demanda por autonomia e pela possibilidade de autodeterminar-se em suas diferentes escolhas e atuações existenciais.

Para tanto, utiliza-se de pesquisa bibliográfica e do método de abordagem hipotético-dedutivo e de interpretação sociológico, buscando compreender as repercussões do fenômeno abordado e os impactos das características do microtrabalho no exercício da dignidade humana. A partir das opções metodológicas descritas, desenvolvem-se noções conceituais acerca do trabalho em plataformas digitais para, em seguida, apresentar as principais características do microtrabalho e sua relação com o desenvolvimento de inteligência artificial. Por fim, à luz das particularidades do microtrabalho identificadas na literatura científica, articula-se o impacto deste cenário na concretização da dignidade da pessoa humana.

1 A PLATAFORMIZAÇÃO DO TRABALHO: CONCEITUAÇÃO E HETEROGENEIDADE DAS ATIVIDADES LABORATIVAS DIGITAIS

A plataformização do trabalho emerge e se consolida enquanto impacto do avanço das novas tecnologias de informação e comunicação. Tendo em vista que tanto o desenvolvimento tecnológico, que encontra amparo na atividade humana, quanto tal aperfeiçoamento, que repercute nas relações de produção, trabalho e tecnologia, encontram-se imbricados historicamente (Grohmann, 2020). Nesse sentido, fenômenos como a possibilidade de realizar atividades laborativas em espaços diversos do estabelecimento do empregador por meio de instrumentos telemáticos (Fincato; Cracco Neto, 2013) e o célere processamento de informações e interconexões – caracterizador, conforme Castells (1999), da "sociedade em rede" –

ressoam na sociedade do trabalho e passam a viabilizar a execução de diferentes atividades em espaços alternativos, inclusive, em plataformas digitais.

Os contornos do trabalho mediado por plataformas digitais podem ser identificados, conforme aponta Pochmann (2016), a partir de tendências afloradas na década de 1970, como a transnacionalização das relações trabalho, o desenvolvimento de novas tecnologias de informação e comunicação e o crescente excedente de mão de obra, que acaba por ser absorvido por contextos laborais que se encontram ao largo de regulamentação. Em atenção ao cenário nacional e ao pesquisar o gerenciamento dessas atividades, Abílio (2019) também indica que a cooptação de trabalhadores para que atuem perante plataformas digitais relaciona-se a um contexto já profundamente atravessado pelo desemprego, trabalho informal e necessidade constante de obtenção de renda de forma fragmentada e insegura – movimento referido como "viração".

Slee (2016), ao pautar a relação entre a Economia do Compartilhamento e o desenvolvimento de empresas-aplicativo, entende que a operação das plataformas sumariza-se, essencialmente, no fornecimento de conexões entre trabalhadores e demanda. A obtenção de lucro a partir da viabilização do contato entre demanda e trabalho, embora invoque características de modos de produção de mão de obra anteriores (Antunes, 2020) e possa ser relacionada aos contratos de terceirização (Kalil, 2019), é marcada por importantes atualizações na organização e gestão do trabalho. Dentre estas particularidades, registra-se a execução do controle das atividades por meio gerenciamento algorítmico (Möhlmann; Zalmanson, 2017), assim como a identificação, pela literatura científica, de um contexto do trabalho perpassado por assimetrias informacionais (Rosenblat; Stark, 2016) e pelo incremento de desigualdades sociais – fazendo-se uso, de forma predominante, de força de trabalhado racializada e de baixa renda (Van Doorn, 2017).

No que tange às propostas de recepção destas atividades no ordenamento juslaboral, Oliveira, Carelli e Grillo (2020) mapeiam os principais caminhos defendidos na abordagem do trabalho mediado por plataformas digitais, quais sejam: (i) proposição de negativa da incidência de qualquer regulação juslaboral; (ii) enquadramento como nova categoria, do "trabalho dependente", que não se encerra na categoria de trabalho autônomo ou subordinado, mas que exige rol protetivo próprio e alinhado ao texto constitucional, conforme desenvolve Kalil (2019); (iii) como nova relação de trabalho, em sentido similar à solução encontrada aos trabalhadores portuários avulsos, estendendo-se os direitos trabalhistas aos trabalhadores de

plataforma; e (iv) tendo em vista a possibilidade de subordinação telemática prevista no parágrafo único do art. 6º da CLT[3], o tradicional reconhecimento de relação de emprego.

Em que pese o objetivo do presente estudo não se direcione à discussão do enquadramento jurídico dos trabalhadores submetidos ao trabalho plataformizado, destaca-se a frequente referência à necessidade de promoção do trabalho decente neste cenário. Nesta senda, Barzotto e Lanner (2019), em análise sob o prisma da fraternidade, enfatizam a vulnerabilidade do trabalho em plataforma e defendem a necessidade de inclusão deste grupo no guarda-chuva dos direitos fundamentais.

Em decorrência dos desafios impostos na tutela de trabalhadores que desempenham atividades perante plataformas digitais, Adam-Prassl (2019, p. 29) destaca a necessidade de reformulações conceituais para abarcar as tendências deste novo mundo laboral, especialmente no que diz respeito às noções de gestão e controle do trabalho, afirmando que "*it will be crucial to adapt and develop our received legal notions of control to include a much broader range of instructions and control*"[4]. Em outro estudo, o pesquisador aponta que o núcleo comum da operação destas atividades encontra-se na disponibilidade de uma multidão de pessoas interessadas em prestar serviços – viabilizando a manutenção da competitividade entre trabalhadores, os baixos preços pagos pelas tarefas executadas e o auferimento de lucro pela empresa responsável pela gestão da plataforma digital (Prassl; Risak, 2016).

O capitalismo de plataforma – concebido por Srnicek (2016) como um fenômeno vinculado à utilização de estruturas digitais responsáveis por proporcionar o contato entre diferentes grupos –, importa ressaltar, extrapola as atividades gerenciadas por empresas como Uber e Rappi. Em consideração às diferentes compreensões acerca do trabalho plataformizado e amparando-se na conceituação recorrente na literatura científica acerca do labor nesse contexto, Grohmann (2020, p. 102-3) condensa ensinamentos e propõe a discriminação das seguintes atividades abarcadas no escopo do trabalho plataformizado:

---

3 "Parágrafo único. Os meios telemáticos e informatizados de comando, controle e supervisão se equiparam, para fins de subordinação jurídica, aos meios pessoais e diretos de comando, controle e supervisão do trabalho alheio." (BRASIL. *Consolidação das Leis do Trabalho (CLT)*. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del5452.htm. Acesso em: 14 set. 2021)

4 Em livre versão: "Será crucial adaptar e desenvolver nossas noções jurídicas de controle para incluir uma gama muito mais ampla do que se entende por instruções e controle".

> a) plataformas que requerem o trabalhador em uma localização específica (como iFood, Rappi, Uber e Deliveroo), as mais conhecidas do cenário do trabalho digital; b) plataformas de microtrabalho ou *crowdwork* (como Amazon Mechanical Turk, PiniOn, Microworkers), marcadas principalmente pelo trabalho de treinar dados para a chamada "inteligência artificial"; c) plataformas *freelance, cloudwork* ou macrotrabalho (como GetNinjas, We Do Logos, Freelancer.com, iPrestador, Fiverr, 99designs), que reúnem tarefas desde pintura e passeio com animais até *design* e programação.

Em sentido similar, De Stefano (2016, p. 2-3), em relatório para a OIT, afirma que o termo "*gig economy*" – que Abílio (2019) registra como termo "guarda-chuva" para aludir ao trabalho mediado por empresas de plataformas digitais, devendo ser recepcionado com cautela e em atenção a questões socioeconômicas próprias do cenário nacional – contempla tanto o "*crowdwork*" quanto o "*work on demand*". O primeiro engloba trabalhos mediados por plataformas digitais, mas realizados *online*, enquanto o segundo, por sua vez, diz respeito a atividades tradicionais, como entregas e caronas, mediadas por plataformas:

> Crowdwork is work that is executed through online platforms that put in contact an indefinite number of organisations, businesses and individuals through the internet, potentially allowing connecting clients and workers on a global basis. The nature of the tasks performed on crowdwork *platforms may vary considerably. [...]*
>
> *In "work on-demand via apps", jobs related to traditional working activities such as transport, cleaning and running errands, but also forms of clerical work, are offered and assigned through mobile apps.*[5]

Portanto, embora se reconheça a viabilidade do exame em conjunto do trabalho plataformizado, haja vista as suas características comuns – como o manejo de meios tecnológicos para a distribuição de tarefas que serão desempenhadas por uma força de trabalho escalonável (De Stefano, 2017) –, importa registrar a existência de particularidades em cada contexto e a importância de pesquisas comprometidas com estas realidades (Sundararajan, 2016). Em atenção a essas especificidades, a presente pes-

5 Em livre versão: "*Crowdwork* é o trabalho executado através de plataformas *online* que colocam em contato um número indefinido de organizações, empresas e indivíduos através da internet, permitindo potencialmente ligar clientes e trabalhadores numa base global. A natureza das tarefas realizadas em plataformas de *crowdwork* pode variar consideravelmente [...] No trabalho '*on-demand*' via aplicativos, demandas relacionadas às atividades de trabalho tradicionais, como transporte, limpeza e realização de tarefas, e também formas de trabalho administrativo, são oferecidas e atribuídas por meio de aplicativos digitais".

quisa volta-se à investigação do microtrabalho, contemplado pela categoria "*crowdwork*", quando direcionado à alimentação e desenvolvimento de *machine learning*.

2 MICROTRABALHO E INTELIGÊNCIA ARTIFICIAL: O LABOR HUMANO EM MEIO À APRENDIZAGEM DE MÁQUINA

Com frequência, a articulação entre aspectos da sociedade do trabalho e inteligência artificial enfatizam a crescente substituição de mão de obra humana por máquinas com capacidade de aprendizagem e de desenvolvimento de tarefas complexas. No que tange à discussão sobre uma possível extinção do trabalho humano, Harari (2018) identifica tendência a um cenário em que as máquinas serão capazes de reproduzir até mesmo o que costuma se referir como "intuição humana", a partir da compreensão de mecanismos bioquímicos e do reconhecimento de padrões. O autor, além de apontar para um futuro atravessado por constantes mudanças e necessidades de adaptação, defende que as atividades de cuidado e altamente especializadas devem se mostrar como as mais resistentes à substituição pela inteligência artificial.

Todavia, em que pese a inclinação para maior demanda de mão de obra especializada, os processos de desenvolvimento de inteligência artificial também são sustentados por trabalho humano pouco qualificado e à margem de regulação juslaboral. Nesse sentido, destacam-se as atividades desempenhadas por microtrabalhores, que consistem, conforme Rosenfield e Mossi (2020, p. 742-3), em "tarefas subindividuais em unidades muito pequenas, resultando em uma remuneração muito baixa, que são lançadas à 'multidão' por intermédio das plataformas para serem executadas".

De acordo com relatório da OIT, as plataformas de microtrabalho despontam com base em iniciativas desenvolvidas pela Amazon a partir de 2005, quando fundada a Amazon Mechanical Turk (AMT). Embora em sua origem estivesse voltada a atender demandas internas da empresa, a AMT passou a abrir a execução de atividades ao público externo e a ofertar a realização de HITs – *Human Intelligence Tasks* (Berg *et al.*, 2018). Na atualidade, a realização de microtarefas extrapolou os limites de Amazon, contando com diversas novas plataformas, como a Microworkers, Clickworker, CrowdFlowe, Prolific e Universal Human Relevance System (UHRS) (Rosenfield; Mossi, 2020; Tubaro Casilli; Coville, 2020).

Entende-se, nesse contexto, o microtrabalho como um complexo de pequenas tarefas fragmentadas, que são desempenhadas de forma remota por um grande número de pessoas – uma multidão (Tubaro; Casilli, 2019). Esta mão de obra digital "*micropaid*" costuma envolver tarefas pequenas, repetitivas e que, geralmente, não demandam qualquer conhecimento especializado (Casilli, 2017). Desta forma, Casilli (2021, p. 29) esclarece que a conjuntura do microtrabalho não enseja a configuração de um emprego *per se*, mas que se trata de trabalhos não convencionais e fora do padrão:

> Eles são, na verdade, trabalhos por peça. Essas pessoas são pagas com apenas alguns centavos ou poucos euros para executar tarefas bem fragmentadas, que, via de regra, são extremamente úteis para treinar e calibrar algoritmos e inteligência artificial.

Este trabalho, portanto, apresenta-se como um resquício – pouco ou nada visualizado – de demanda humana em meio aos processos de desenvolvimento de *machine learning*, subsidiando a revolução da inteligência artificial mediante a submissão de uma multidão de trabalhadores, com poucas condições de selecionar suas atividades profissionais no mercado de trabalho existente, a condições precárias e incertas de labor (Gray; Suri, 2017). Registra-se que essas atividades se diferenciam do trabalho de *freelancers* desempenhado em nuvem, na medida em que o *freelancer* direciona-se à realização de trabalhos criativos, por profissionais qualificados, que se envolvem em um projeto inteiro ou longo, enquanto o microtrabalho é desempenhado de maneira apartada da totalidade da atividade e caracteriza-se pela simplicidade e curto lapso temporal entre uma *task* e outra (Tubaro *et al.*, 2020).

Acrescenta-se que, embora apresente a fragmentação da atividade como elemento predominante, o microtrabalho não se confunde com um mero "taylorismo digital". Ainda que a parcialização da atividade configure como elemento central do taylorismo, o trabalhador neste modelo integrava a um todo maior conhecido, havendo tanto assalariamento como relativa segurança, especialmente vinculada à atuação coletiva e movimentos de agremiação naquela conjuntura. No microtrabalho, contudo, a parcialização da atividade atinge novo patamar, em que o microtrabalhador, com frequência, desconhece o todo ao qual a tarefa executada integra e encontra-se exposto a uma nova temporalidade e especialidade de trabalho, com pouco ou nenhuma amparo social e coletivo (Rosenfield; Mossi, 2020, p. 748).

Ao investigar o microtrabalho em cenário nacional, Braz (2021, p. 144-7) discrimina as cinco principais formas como o microtrabalhar se expressa no Brasil, quais sejam: (i) para treinamento e produção de dados, de forma muito vinculada a plataformas globais de inteligência artificial; (ii) para realização de pesquisas de mercado; (iii) para impulsionamento de redes sociais, também conhecidas como "fazendas de cliques"[6]; (iv) para pequenos serviços de realização de depoimentos, divulgações de textos e contas, dentre outros; e (v) para testes de usabilidade remota, visando a testagem e levantamento de informações sobre novos produtos.

Ainda que se reconheça que o microtrabalho contemple uma variedade de pequenas tarefas realizadas em *crowdwork*, a presente pesquisa preocupa-se com o microtrabalho desempenhado para fins de alimentação e treinamento de inteligência artificial. Neste aspecto, Tubaro e Casilli (2019) esclarecem que as atividades realizadas por microtrabalhadores ajudam a subsidiar os processos vinculados à inteligência artificial gerando dados, assim como solucionando atividades que as máquinas não executam de forma eficaz. Desta forma, as microtarefas se encontram a serviço da superação dos "*gaps*" da inteligência artificial (Willow Garage, 2009).

Dentre as atividades exercidas neste modelo, em que plataformas ofertam tarefas em um mercado *online* para empresas que atuam no aperfeiçoamento de inteligência artificial, constatam-se incumbências como realizar anotações em imagens e gravar áudios com solicitações usualmente realizadas para assistentes virtuais (Berg *et al.*, 2018). Tubaro (2020, p. 38) registra algumas expressões deste trabalho direcionado a uma multidão de trabalhadores sem contrato formal e por meio de dispositivos digitais:

> *The extraordinary successes of present-day artificial intelligence (AI) rest on the "microwork" of a multitude of real men and women. They tag objects in images, transcribe commercial receipts, translate bits of text, and record their voice while reading aloud short sentences. Simple and repetitive, these tasks generally require low qualifications and are paid as little as a few cents.*[7]

6 Registra-se que, nas "*click farms*", a atividade dos microtrabalhadores consiste na contínua realização de reações a publicações e páginas *online*, tendo um profundo impacto sociopolítico (Kuek *et al.*, 2015).

7 Em livre versão: "O extraordinário sucesso da inteligência artificial (IA) atual repousa no 'microtrabalho' de uma multidão de homens e mulheres reais. Eles marcam objetos em imagens, transcrevem recibos comerciais, traduzem trechos de texto e gravam sua voz enquanto leem frases curtas em voz alta. Simples e repetitivas, essas tarefas geralmente exigem baixas qualificações e são remuneradas apenas com alguns centavos".

Ainda no que tange às atividades desempenhadas para contribuir ao desenvolvimento de inteligência artificial, observa-se que, na indústria automotiva, as principais funções exercidas por microtrabalhadores dizem respeito à classificação de imagens, detecção ou marcação de objetos, identificação de pontos de referência e segmentação semântica. Assim, a título de exemplo, as tarefas demandadas aos trabalhadores subsidiam a aprendizagem da máquina auxiliando na identificação de elementos como animais atravessando as ruas e bicicletas (Tubaro; Casilli, 2019).

Ao sistematizar maneiras como o microtrabalho ampara a inteligência artificial, Tubaro *et al*. (2020) apresentam três eixos que traduzem esta vinculação em diferentes momentos da criação e aperfeiçoamento de *machine learning*, quais sejam: (i) na preparação da IA, (ii) na personificação da IA e (iii) na verificação da IA. A preparação da inteligência artificial encontra-se subdividida em duas fases – geração de dados e anotações de dados. A geração de dados pode ser ilustrada pela gravação de áudios com diferentes sotaques; na anotação de dados, por seu turno, encontram-se demandas como a categorização de tópicos de conversas e a determinação de emoções humanas relacionadas a alguma declaração.

A personificação da inteligência artificial, que Irani (2015) descreve como a parte "artificial" da própria "inteligência artificial", consiste em medida paliativa que diz respeito tanto à superação de atividades que a máquina ainda não consegue desempenhar com eficiência quanto à busca, por parte das empresas, de ocultamento da necessidade de mão de obra humana, visando atrair acionistas e sustentar a imagem de autonomia da máquina divulgada publicamente. Por conseguinte, é identificada especialmente quando *startups* que se apresentam como desenvolvedoras de inteligência artificial utilizam-se, para redução de custos, do labor de microtrabalhadores contratados mediante plataformas digitais (Tubaro *et al.*, 2020, p. 7):

> *Costs go further up if accuracy of results is soughts [...]. Under pressure to perform, companies may find it cheaper to just leave aside cutting-edge technology, fragment the work into micro-tasks and sub-contract them to low-paid workers through platforms.*[8]

8 Em livre versão: "Os custos aumentam ainda mais se a precisão dos resultados for buscada [...]. Sob pressão para apresentar um bom desempenho, as empresas podem achar mais barato simplesmente deixar de lado a tecnologia de ponta, fragmentar o trabalho em microtarefas e subcontratá-las com trabalhadores de baixa remuneração por meio de plataformas".

Constata-se, ainda, o uso dos serviços de plataformas de microtrabalho para verificação da inteligência artificial. Isto é, demandas endereçadas à confirmação e correção da atuação da inteligência artificial – como analisar se o assistente virtual entendeu corretamente uma determinada declaração e teve sua operação em consonância com as expectativas do usuário (Tubaro *et al.*, 2020).

Em atenção às diferentes maneiras como o microtrabalho é demandado pela inteligência artificial, observa-se que a imagem ideal de plena autonomia da máquina em relação ao humano, que frequentemente acompanha iniciativas relacionadas à *machine learning*, ainda não se sustenta (Irani, 2015) e apresenta larga contradição quando se investiga a realidade das microtarefas e dos trabalhadores que as desempenham. Verifica-se, à vista disso, que estes resquícios de atividades humanas pouco qualificadas atravessam a indústria da tecnologia, apresentando-se enquanto uma demanda estrutural da indústria de *machine learning* (Gray; Suri, 2019). Em mesmo sentido, pesquisas desenvolvidas por Tubaro *et al.* (2020, p. 4) apontam à continuidade do vínculo entre microtarefas e inteligência artificial:

> *Our other expectation is that micro-work is a structural rather than a temporary input to AI production. While some technology enthusiasts believe that data generation and annotation tasks will ultimately be fully automated, the "heteromation" paradigm (Ekbia and Nardi, 2017) implies that some essential tasks will always be directed to humans as indispensable though hidden providers. [...] Our review of machine learning techniques corroborates this line of thought: their huge and growing data needs will keep demand for micro-work high in the foreseeable future.*[9]

O "toque humano" do microtrabalho manifesta-se, portanto, como uma contribuição que diz respeito à estrutura da inteligência artificial, tendo em consideração que seus processos envolvem uma grande quantidade de dados, e não meramente temporária. Isto é, na conjuntura de desenvolvimento da inteligência artificial baseada em aprendizagem estatística, a demanda por dados não atingirá um estado estacionário, devendo ser continuamente adaptado às mudanças para gerar resultados mais precisos

9 Em livre versão: "Nossa outra expectativa é que o microtrabalho seja mais estrutural do que temporário para a produção de IA. Embora alguns entusiastas da tecnologia acreditem que as tarefas de geração e anotação de dados serão, em última análise, totalmente automatizadas, o paradigma de 'heteromação' (Ekbia e Nardi, 2017) implica que algumas tarefas essenciais sempre serão direcionadas aos humanos, enquanto provedores indispensáveis, embora ocultos. [...] Nossa revisão da aprendizagem de máquina corrobora para esta linha de pensamento: as enormes e crescentes necessidades de dados manterão alta demanda por microtrabalho em um futuro previsível".

e lucrativos (Tubaro *et al.*, 2020). Diante do cenário descrito pela literatura científica, importa investigar as condições em que as atividades laborativas de microtrabalhadores são exercidas, especialmente à luz do princípio e valor constitucional da dignidade da pessoa humana.

3 A FRUIÇÃO DA DIGNIDADE DA PESSOA HUMANA NA CONJUNTURA DO MICROTRABALHO: A INVISIBILIZAÇÃO E INSTRUMENTALIZAÇÃO DO ELEMENTO HUMANO

Em decorrência das particularidades do contexto do microtrabalho digital, o claro mapeamento dos trabalhadores que se encontram vinculados a plataformas que fornecerem microtarefas para alimentação de inteligência artificial é dificultado. Dentre os obstáculos reconhecidos, encontra-se a existência de diferentes níveis de engajamento com a atividade, haja vista que as microtarefas são executadas tanto para fins de obtenção de renda principal de trabalhadores quanto, por outro lado, como atividade exercida em curtos lapsos temporais livres do microtrabalhador (Berg *et al.*, 2018).

Em pesquisa investigando as dificuldades metodológicas para contagem de microtrabalhadores, Tubaro, Ludec e Casilli (2020) apontam, além dos desafios inerentes à diversidade de engajamento e do papel desta renda na organização financeira de cada microtrabalhador, também a ausência de âncora geográfica – coexistindo trabalhadores de diferentes nacionalidades – e o uso concomitante de grande variedade de plataformas digitais. Ainda, os autores indicam as dificuldades decorrentes da subnotificação da atividade e da ausência de percepção, pelo próprio microtrabalhador, do caráter laborativo das tarefas:

> *Micro-work is more likely than other platform-based earning activities to be under-reported in general social surveys. The simple, short and sometimes "gamified" nature of micro-tasks misleads workers who often fail even to recognise them as labour.*[10]

Em que pese a existência de adversidades para a elaboração de uma cartografia precisa dos espaços e condições do microtrabalho no cenário do labor digital, faz-se possível identificar um contexto caracterizado pela insegurança e falta de reconhecimento. Nesse sentido, relatório promovido pela

10 Em livre versão: "É mais provável que o microtrabalho seja subnotificado em pesquisas sociais em geral do que outras atividades lucrativas baseadas em plataforma. A natureza simples, curta e às vezes 'gamificada' das microtarefas engana os trabalhadores que muitas vezes nem conseguem reconhecer estas atividades como trabalho".

OIT (Berg *et al.*, 2018) indica preocupação com a ausência de regulamentação da atividade, que tem suas condições determinadas unilateralmente pelas plataformas digitais – não havendo, muitas vezes, sequer garantia de adimplemento pela tarefa realizada, especialmente quando a empresa que contratou o serviço entende que a sua execução não alcançou o merecimento de contraprestação.

Além de apontar para uma situação de precarização ainda mais acentuada em contexto nacional, Braz (2021) salienta a assimetria da relação estabelecida entre os trabalhadores e as plataformas em uma relação em que o labor é desempenhado sem margens para negociações e proteção social. O autor atenta, ainda, aos impactos decorrentes da ausência de compreensão da totalidade e finalidade de sua tarefa por parte trabalhador, culminando tanto em uma maior dificuldade na construção de laços sociolaborais sólidos, quanto induzindo à redução da atividade humana a mero funcionamento procedimental, dissociada de sentido (Braz, 2021).

A invisibilização da necessidade de microtrabalhadores, com o objetivo de manter uma aparência que gere maiores investimentos por acionistas e aceitação por consumidores, implica uma larga fenda entre a imagem difundida do aperfeiçoamento de processos de *machine learning* e a realidade laboral dos trabalhadores que executam microtarefas (Irani, 2015, p. 9):

> *Within the private sector, automatic management also serves a symbolic purpose with financial consequence. Microwork companies attract more generous investment terms when investors perceive them as technology companies rather than labor companies.*[11]

Em entrevistas com responsáveis pela gestão de microtarefas, Irani (2015) identifica que os problemas dos trabalhadores são tratados pela gestão das plataformas digitais como problemas computacionais, sob a égide de um eixo estritamente direcionado à precisão, velocidade e escalabilidade. Desta forma, os trabalhadores têm a sua singularidade silenciada frente à imposição de um tratamento como se máquinas fossem.

Além da promoção de medidas para manter o microtrabalhador longe do olhar público em geral e da agenda política, verifica-se a acentuada confusão dos trabalhadores que executam as microtarefas acerca do propósito

11 Em livre versão: "Dentro do setor privado, a gestão automática também serve a um propósito simbólico com consequências financeiras. As empresas de microtrabalho atraem investimentos mais generosos quando os investidores as percebem como empresas de tecnologia em vez de empresas de trabalho".

de sua atividade, obstaculizando a integração social e a autonomia destes grupos. Em pesquisa de Tubaro *et al.* (2020, p. 6-7), um trabalhador que desempenha estas pequenas tarefas, ao ser entrevistado, afirma "*then I think to myself: if it's there, it must be useful to someone, for something, but... Why, I don't know*"[12].

Com resultados similares, pesquisa desenvolvida por Grohmann e Araújo (2021) buscou investigar a percepção dos próprios trabalhadores acerca das atividades desempenhadas junto às plataformas que ofertam microtarefas à inteligência artificial, e localizaram postagens de microtrabalhadores questionando-se acerca de sua identidade laboral. Os autores apontam que as noções de dados e inteligência artificial são raramente mencionadas entre os trabalhadores – sendo a inteligência artificial mencionada espontaneamente em apenas uma entrevista –, explicitando que muitos trabalhadores não se percebem como parte da cadeia de aperfeiçoamento da aprendizagem de máquinas.

A significativa assimetria informacional entre trabalhadores e plataformas digitais, que culmina na carência de compreensão do microtrabalhador sobre o todo em que sua atividade se insere – e, com frequência, inclusive na incompreensão do caráter laboral de sua atuação junto às plataformas de microtarefas –, implica um facilitador para a continuidade do trabalho em condições precárias. Ademais, a invisibilidade, confusão e solidão no cenário do microtrabalho importam em prejuízos na formação de um elo social protetivo, conforme aduzem Rosenfield e Mossi (2020, p. 749):

> Em última instância, isso contribui ao enfraquecimento do elo social e dos mecanismos de suporte societários, pois não há, pretensamente, humanos envolvidos. A quem reclamar e reivindicar? É uma organização ímpar no que se refere à total independência do trabalhador em relação àquele que demanda seu trabalho, ao fato de o trabalhador entregar o produto do seu trabalho podendo não ser aceito nem ser pago, à ausência total de compromisso do cliente e do operador da plataforma em relação ao trabalhador. Este está inserido precariamente na esfera do mundo do trabalho, desenlaçado e sem qualquer vínculo ou proteção. É um trabalhador invisível, "solto" no processo de trabalho.

Diante desse cenário, a preocupação com as possibilidades para concretização do princípio e valor constitucional da dignidade da pessoa huma-

12 Em livre versão: "Então eu penso comigo mesmo: se está aí, deve ser útil para alguém, para alguma coisa, mas... por quê, eu não sei".

na merece atenção. Conforme ensina Sarlet (2019), a dignidade é qualidade intrínseca e irrenunciável da pessoa humana. Por conseguinte, a dignidade não pode ser concedida ou retirada da pessoa humana, mas deve ser protegida e respeitada. Embora inexista conduta patronal capaz de retirar a dignidade do trabalhador humano, ante o seu caráter inerente, reconhece-se a possibilidade de manutenção de condutas e condições laborais que ferem este princípio e valor fundante do ordenamento jurídico pátrio, positivado no inciso III do art. 1º da Constituição Federal de 1988.

Destaca-se que se encontra consagrada no ordenamento constitucional a vedação de práticas que submetam as pessoas a situações degradantes ou que violem a dignidade humana, inerente a todo indivíduo (Siqueira, 2016). Ainda, atento ao risco de banalização do valor-guia da dignidade da pessoa humana, Santos Junior (2010) aduz que o impedimento a sua fruição – haja vista que não há falar em pessoa humana sem dignidade – deve ser analisado no caso em concreto, ratificando a perspectiva da dignidade como intrinsecamente vinculada à vedação da redução do humano à condição de objeto.

A dignidade humana – que não se encerra em um conceito fixista e permanece em constante desenvolvimento – vincula-se, portanto, à doutrina kantiana ao anunciar impedimento de instrumentalização do ser humano – isto é, sujeitá-lo à condição que negue a sua distinção enquanto humano. Tal valor, contudo, não impede que o humano preste serviços em favor de objetivos de terceiros, mas sim que reste encerrado em uma relação que culmine em sua coisificação (Sarlet, 2019, p. 61):

> [...] vale registrar, ainda, que mesmo Kant nunca afirmou que o homem, num certo sentido, não possa ser "instrumentalizado" de tal sorte que venha a servir, espontaneamente e sem que com isto venha a ser degradado na sua condição humana, à realização de fins de terceiros, como ocorre, de certo modo, com todo aquele que presta um serviço a outro. Com efeito, Kant refere expressamente que o homem constitui um fim em si mesmo e não pode servir "simplesmente como meio para o uso arbitrário desta ou daquela vontade".

Com amparo na doutrina de Dieter Grimm, Sarlet (2019) ensina que a dignidade humana resta violada quando atenta à autonomia da pessoa de decidir acerca de seus projetos existenciais (Sarlet, 2019). A autonomia, portanto, é elemento fundante da dignidade (Weber, 2013).

Acrescenta-se que a concretização do princípio da dignidade da pessoa humana entrelaça-se, também, à realização de direitos fundamentais formalmente registrados na Constituição Federal ou decorrentes da abertura material do catálogo constitucional (Sarlet, 2007). Imprescindível, portanto, alcançar a realização substantiva das garantias sociais fundamentais (Streck, 2002), vez que a efetivação de direitos nas relações de trabalho consolida-se enquanto fundamental na afirmação da cidadania social e ampara a manutenção da democracia (Delgado, 2007). Outrossim, considerando as reformulações decorrentes do avanço tecnológico, é salutar a reflexão acerca da concretização do princípio da dignidade humana do trabalhador frente às repercussões das novas tecnologias nesta seara (Stürmer; Coimbra, 2018).

A partir da compreensão de dignidade profundamente relacionada à autonomia do humano de autodeterminar-se e gerir suas práticas existenciais, o cenário de desenvolvimento e aprimoramento de inteligência artificial, ao se utilizar do desempenho de atividades de microtrabalhadores, é potencial ensejador de condições violadoras da dignidade da pessoa humana. Na medida em que invisibiliza todo o grupo de trabalhadores que executam microtarefas e sustenta acentuada assimetria informacional, o cenário do microtrabalho impede que os trabalhadores assimilem a finalidade da sua atividade e, até mesmo, sejam reconhecidos perante consumidores, acionistas e pares.

Nesse sentido, ao examinar as recomendações da Organização Internacional do Trabalho (OIT), Rosenfield e Mossi (2020, p. 754) indicam que "as relações trilaterais entre plataforma, clientes e trabalhadores, por serem abusivas, colocam em jogo o respeito à dignidade do microtrabalhador". As pesquisadoras enfatizam que, para além de obtenção de melhores condições de trabalho em âmbito de distribuição – como remuneração –, o cenário do microtrabalho gera preocupante demanda por reconhecimento.

Constatam-se, desta forma, violências que se expressam no âmbito moral e de relações intersubjetivas de reconhecimento, sendo anteriores às restrições de dimensões operacionalizáveis. Portanto, além de demandas de representação coletiva e de distribuição – como a garantia de percebimento de remuneração pela atividade realizada e de jornada razoável –, a atuação deste grupo de trabalhadores na alimentação de sistemas de inteligência artificial vincula-se essencialmente a uma dimensão moral e que diz respeito à dignidade do trabalhador (Rosenfield; Mossi, 2020, p. 758):

> [...] a postura desrespeitosa (um mau tratamento, ou um tratamento abusivo) em relação aos microtrabalhadores impede que sejam reconhecidos e se reconheçam como sujeitos dignos de tomar parte na vida em sociedade, colocando em questão o valor moral desses sujeitos e de suas capacidades e realizações.

Diante desta conjuntura, percebe-se que o predominante cenário de desenvolvimento de inteligência artificial baseada em aprendizagem estatística, que faz uso de mão de obra humana tanto para gerar e anotar dados como para verificar a correção da atuação da máquina e, até mesmo, falsear um funcionamento supostamente autônomo da inteligência artificial, culmina em condições de trabalho que prejudicam a fruição da dignidade humana garantida em texto constitucional.

À luz das preocupações identificadas em literatura científica, sublinha-se que tanto a camuflagem da demanda por atividades humanas na realização de pequenas tarefas, vinculada à tentativa de manutenção de uma imagem de atuação autônoma da máquina perante consumidores e acionistas, quanto a tratativa deste labor a partir de uma perspectiva meramente computacional – conforme identificado em estudo de Irani (2015) – são aspectos do contexto do microtrabalho que implicam a instrumentalização do humano.

Importa notar que o trabalho é dimensão de afirmação da identidade do sujeito, perante si e perante pares, sendo fator de inserção social e sentido (Bendassolli; Soboll, 2011). O manejo do labor humano enquanto máquina e a marcada tendência a velar a mão de obra humana no cenário da inteligência artificial são geradores, por consequência, de prejuízos à dimensão social dos trabalhadores. Ademais, a carência de percepção de sua atividade enquanto ocupação laborativa – ainda que não se discuta, no presente artigo, a pertinência de classificação aos modelos tradicionais de emprego – atrofia as possibilidades de articulação com pares e de negociação para alcance de melhores condições laborais.

Por consequência, a acentuada assimetria informacional conecta-se com a manutenção de um trabalho invisível e dissociado de sentido social, sendo preocupação recorrente em relatório da Organização Internacional do Trabalho a promoção de medidas capazes de impulsionar maior simetria nas relações e o reconhecimento das atividades laborais caracterizadas como microtrabalho. Destaca-se, nesta linha, a recomendação de informar

aos microtrabalhadores os objetivos de seu trabalho e a identidade dos clientes que se utilizam das tarefas cumpridas (Berg, 2018).

Portanto, embora os processos de *machine learning* estejam endereçados ao alcance de aprendizagem semelhante ou superior à de humanos, gerando grande preocupação no que tange às taxas de ocupabilidade de mão de obra humana em um futuro próximo, denota-se que a indústria tecnológica tende a velar a necessidade de execução de microtarefas por humanos. Este mascaramento, associado à tratativa do elemento humano como se máquina fosse, enseja a coisificação de trabalhadores que, com frequência, sequer percebem-se enquanto parte da indústria a que pertencem.

A promoção de condições para a devida fruição da dignidade humana, em conformidade com os preceitos constitucionais, tendo como pressuposto o exercício de autonomia e a possibilidade de autodeterminar-se em suas escolhas e existência, envolve a aplicação de medidas que desembaracem a visibilidade deste grupo perante a sociedade e que viabilizem que o trabalhador perceba a realização de suas tarefas enquanto atividades laborativas dotadas de sentido, que alimentam a indústria da inteligência artificial. A dignidade, na conjuntura do microtrabalho, depende do reconhecimento do caráter humano – e de suas respectivas demandas sociais e simbólicas – dos sujeitos que desempenham microtarefas, para a superação de um cenário em que máquina promove o achatamento e a incompreensão do trabalhador, ainda, humano.

CONSIDERAÇÕES FINAIS

A partir da preocupação com os prejuízos ao exercício da dignidade da pessoa humana no contexto do microtrabalho, especialmente no que tange às microtarefas executadas para alimentação e aperfeiçoamento de processos de aprendizagem de máquina baseada em estatísticas, identificou-se a necessidade de compreender aspectos deste cenário capazes de ensejar a instrumentalização do humano na indústria da inteligência artificial. Por meio de revisão de literatura científica e doutrinária, fez-se possível identificar que a plataformização do trabalho, fenômeno que desponta a partir da década de 70 e concretiza-se frente à possibilidade de plataformas digitais gerirem e mediarem a relação entre trabalhadores e agentes interessados na prestação de serviços, é marcado pela pluralidade.

Desta forma, para além das conhecidas atividades "*on-demand*", como as vinculadas a aplicativos de entrega e caronas, optou-se por privi-

legiar o estudo do trabalho em "*crowdwork*", especialmente do microtrabalho. Caracterizado pelo desempenho de pequenas tarefas, de natureza variada, o microtrabalho é responsável pela alimentação do desenvolvimento e aperfeiçoamento de inteligência artificial baseada em aprendizagem estatística. Tendo em vista que tais processos demandam grande quantidade de dados, assim como anotações, verificações e preenchimentos de "*gaps*" ainda não superados pela *machine learning*, verifica-se que a demanda de trabalho humano pouco qualificado permanece entrelaçado à indústria da inteligência artificial.

Além disso, constata-se que a necessidade de pequenas atividades desempenhadas por humanos é, com frequência, ocultada pelos desenvolvedores de inteligência artificial. As relações estabelecidas nesta seara, ademais, são marcadas por assimetrias informacionais e pelo manejo da atividade humana como mero elemento computacional, culminando em uma conjuntura em que o microtrabalhador guarda pouca ou nenhuma compreensão acerca da finalidade da sua atividade – muitas vezes, sequer entendendo que as microtarefas se tratam de atividades laborativas, dotadas de sentido institucional e social.

Diante disso, identifica-se acentuada ameaça à fruição da dignidade humana na conjuntura do labor perante plataformas digitais de microtrabalho. A promoção de medidas que coíbam a instrumentalização do humano no desenvolvimento de inteligência artificial perpassa, por consequência, a necessidade de estratégias que gerem maior transparência acerca da demanda de atividade humana na verificação e alimentação da inteligência artificial.

Tendo em vista que o pleno exercício da dignidade reclama por autonomia e pela possibilidade de autodeterminar-se em suas escolhas, inclusive no que se refere à ocupação laborativa, o fomento de estratégias endereçadas à redução da invisibilidade e carência de reconhecimento social e institucional destes grupos de trabalhadores se faz essencial. Por fim, a partir dos resultados obtidos a partir da pesquisa bibliográfica desenvolvida, aponta-se para a necessidade de pesquisas futuras que investiguem e sugiram alternativas de políticas públicas e de atuação de agentes sociais voltadas ao reconhecimento sociocultural dos trabalhadores vinculados às plataformas de microtrabalho.

REFERÊNCIAS

ABILIO, Ludmila Costhek. Uberização: do empreendedorismo para o autogerenciamento subordinado. *Psicoperspectivas*, v. 18, n. 13, 2019. Disponível em: https://www.cesit.net.br/uberizacao-do-empreendedorismo-para-o-autogerenciamento-subordinado/. Acesso em: 13 set. 2021.

ADAMS-PRASSL, Jeremias. What if your boss was an algorithm? Economic incentives, legal challenges, and the rise of artificial intelligence at work. *Comparative Labor Law & Policy Journal 123*, v. 41, n. 1, 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3661151. Acesso em: 14 set. 2021.

_____; RISAK, Martin. Uber, TaskRabbit, & CO: platforms as employers? Rethinking the legal analysis of crowdwork. *Oxford Legal Studies Research Paper*, n. 8, 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2733003. Acesso em: 14 set. 2021.

BARZOTTO, Luciane Cardoso; LANNER, Maíra Brecht. Declaração do centenário da Organização Internacional do Trabalho e a proteção do trabalhador digital no paradigma da fraternidade. *E-Revista Internacional de la Protección Social*, v. IV, n. 2, p. 125-143, 2019.

BENDASSOLLI, Pedro F.; SOBOLL, Lis Andrea. *Clínicas do trabalho*. São Paulo: Atlas, 2011.

BERG, Janine; FURRER, Marianne; HARMON, Ellie; RANI, Uma; SILBERMAN, Six. *Digital labour platforms and the future of work*: towards decent work in the on-line world. Geneva: International Labour Office, 2018. Disponível em: https://www.ilo.org/global/publications/books/WCMS_645337/lang--en/index.htm. Acesso em: 14 set. 2021.

BRASIL. *Consolidação das Leis do Trabalho (CLT)*. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del5452.htm. Acesso em: 14 set. 2021.

_____. *Constituição da República Federativa do Brasil de 1988*. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 22 set. 2021.

BRAZ, Matheus Viana. Heteromação e microtrabalho no Brasil. *Sociologias*, a. 23, n. 17, p. 134-172, maio/ago. 2021.

CASILLI, Antonio. Digital Labor Studies Go Global: Toward a Digital Decolonial Turn. *International Journal of Communication*, v. 11, 2017, p. 3934-3954. Disponível em: https://ijoc.org/index.php/ijoc/article/view/6349/2149. Acesso em: 17 set. 2021.

_____. O trabalho digital além da uberização. In: GROHMANN, Rafael (Org.). *Os laboratórios do trabalho digital*: entrevistas. 1. ed. São Paulo: Boitempo, 2021.

CASTELLS, Manuel. *A sociedade em rede*. São Paulo: Paz e Terra, v. I, 1999.

DELGADO, Mauricio Godinho. Direitos fundamentais na relação de trabalho. *Revista de Direitos e Garantias Fundamentais*, n. 2, p. 11-39, 2007.

DE STEFANO, Valerio. The rise of the "just-in-time workforce": on-demand work, crowdwork and labour protection in the "gig-economy". *Conditions of Work and Employment Series*, n. 71, Geneva: ILO, 2016.

_____. Labour is not a technology: reasserting the Declaration of Philadelphia in Times of Platform-Work and Gig-Economy. *IUSLabor*, 2, p. 1-16, 2017.

FINCATO, Denise Pires; CRACCO NETO, Heitor Barbieri. Teletrabalho: de Chappe a Nilles. *Revista Justiça do Trabalho*, Porto Alegre, a. 30, n. 358, out. 2013.

GRAY, Mary; SURI, Siddhart. The humans working behind the AI curtain. *Harvard Business Review*, p. 2-5, 9 jan. 2017.

_____; _____. *Ghost work*: how to stop Silicon Valley from building a new global underclasse. Boston: Houghton Mifflin Harcourt, 2019.

GROHMANN, Rafael. Plataformização do trabalho: características e alternativa. In: ANTUNES, Ricardo (Org.). *Uberização, trabalho e indústria 4.0*. São Paulo: Boitempo, 2020.

_____; ARAÚJO, Willian Fernandes. Beyond mechanical turk: the work of Brazilians on global AI platforms. In: VERDEGEM, Peter (Org.). *AI for everyone?* Critical perspectives. London: University of Westminster Press, p. 247-166, 2021. Disponível em: https://www.uwestminsterpress.co.uk/site/books/e/10.16997/book55/. Acesso em: 27 set. 2021.

HARARI, Yuval Noah. *21 lições para o século 21*. 1. ed. São Paulo: Companhia das Letras, 2018.

IRANI, Lilly. The cultural work of microwork. *New Media & Society*, v. 17, n. 5, p. 720-739, 2015. Disponível em: https://escholarship.org/uc/item/2x10h7rs#main. Acesso em: 19 set. 2021.

KALIL, Renan Bernardi. *Capitalismo de plataforma e direito do trabalho*: *crowdwork* e trabalho sob demanda por meio de aplicativos. Tese de Doutorado. Orientador Otavio Pinto e Silva. São Paulo, Universidade de São Paulo, 2019.

KUEK, Siou Chew; PARADI-GUILFORD, Cecilia; FAYOMI, Torks; IMAIZUMI, Saori; IPEIROTIS, Panos. *The global opportunity in online outsourcing*. World Bank, Washington, DC, 2015. Disponível em: https://openknowledge.worldbank.org/handle/10986/22284. Acesso em: 17 set. 2021.

MÖHLMANN, Mareike; ZALMANSON, Lior. Navigating Algorithmic Management and Drivers' Autonomy. *Thirty Eighth International Conference on Information Systems*, South Korea, 2017. Disponível em: https://www.researchgate.net/profile/Mareike_Moehlmann2/publication/319965259_Hands_on_the_wheel_Navigating_algorithmic_management_and_Uber_drivers'_autonomy/

links/59c3eaf845851590b13c8ec2/Hands-on-the-wheel-Navigating-algorithmic-management-and-Uber-drivers-autonomy.pdf. Acesso em: 13 set. 2021.

OLIVEIRA, Murilo Carvalho Sampaio; CARELLI, Rodrigo de Lacerda; GRILLO, Sayonara. Conceito e crítica das plataformas digitais de trabalho. *Revista Direito e Praxis*, Rio de Janeiro, v. 11, n. 4, p. 2628-2629, 2020.

POCHMANN, Marcio. A crise capitalista e os desafios dos trabalhadores. *Cadernos do CEAS*, Salvador, n. 239, p. 698-723, 2016.

ROSENFIELD, Cinara; MOSSI, Thays Wolfarth. Trabalho decente no capitalismo contemporâneo: dignidade e reconhecimento no microtrabalho por plataformas. *Revista Sociedade e Estado*, v. 35, n. 3, p. 741-764, set./dez. 2013.

SANTOS JUNIOR, Rubens Fernando Clamer dos. *A eficácia dos direitos fundamentais dos trabalhadores*. São Paulo: LTr, 2010.

SARLET, Ingo Wolfgang. *A eficácia dos direitos fundamentais*. 8. ed. Porto Alegre: Livraria do Advogado, 2007.

_____. *Dignidade (da pessoa) humana e direitos fundamentais na Constituição Federal de 1988*. 10. ed. Porto Alegre: Livraria do Advogado, 2019.

SIQUEIRA, Rodrigo Espiúca dos Anjos. *Relações de trabalho e direitos fundamentais sociais*. Curitiba: Juruá, 2016.

SRNICEK, Nick. *Platform capitalism*. New York: John Wiley & Sons, 2016.

STRECK, Lenio Luiz. *Jurisdição constitucional e hermenêutica*: uma nova crítica do direito. Porto Alegre: Livraria do Advogado, 2002.

STÜRMER, Gilberto; COIMBRA, Rodrigo. As novas tecnologias e o meio ambiente de trabalho. *Revista de Direito do Trabalho*, São Paulo, v. 192, a. 44, p. 123-148, ago. 2018.

SUNDARARAJAN, Arun. *The sharing economy*: the end of employment and the rise of crowd-based. Cambridge: Mit Press, 2016.

TUBARO, Paola. Whose intelligence is artificial intelligence? *Global Dialogue, International Sociological Association*, 2020, p. 38-39. Disponível em: https://hal.archives-ouvertes.fr/hal-03029735. Acesso em: 14 set. 2021.

_____; CASILLI, Antonio. Micro-work, artificial intelligence and the automotive industry. *Journal of Industrial and Business Economics*, Springer, 2019, p. 1-13. Disponível em: https://hal.archives-ouvertes.fr/hal-02148979. Acesso em: 15 set. 2021.

_____; _____; COVILLE, Marion. The trainer, the verifier, the imitator: Three ways in which human platform workers support artificial intelligence. *Big Data & Society*, p. 1-12, jan./jun. 2020. Disponível em: https://journals.sagepub.com/doi/10.1177/2053951720919776. Acesso em: 17 set. 2021.

VAN DOORN, Niels. Platform Labor: on the gendered and racialized exploitation of low-income servisse work in the "on-demand" economy. *Information, Communication & Society*, v. 20, n. 6, p. 898-914, 2017.

WEBER, Thadeu. *Ética e filosofia do Direito*: autonomia e dignidade da pessoa humana. Petrópolis: Vozes, 2013.

WILLOW GARAGE. *Humans Helping Robots See*, 24 ago. 2009. Disponível em: ros.org/news/2009/08/humans-helping-robots-see.html. Acesso em: 19 set. 2021.

Sobre as autoras:

Denise Pires Fincato | *E-mail:*dpfincato1@gmail.com

Doutora em Direito pela Universidad de Burgos (2001) e Pós-Doutora pela Universidad Complutense de Madrid (2017). Professora-Pesquisadora no Programa de Pós-Graduação da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Coordenadora do Grupo de Pesquisas Novas Tecnologias, Processo e Relações de Trabalho (CNPQ/PUCRS). Advogada e Consultora Trabalhista.

Julise Carolina Lemonje | *E-mail:* juliselemonje@hotmail.com

Mestranda em Direito pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), na área de concentração Fundamentos Constitucionais de Direito Público e Privado. Especializanda em Direito e Processo do Trabalho pela Universidade Federal do Rio Grande do Sul (UFRGS). Advogada e Psicóloga. Bolsista PRO-Stricto PUCRS.

Data de submissão: 29 de setembro de 2021.
Data de aceite: 10 de janeiro de 2022.

Os Sistemas de Armas Autônomas e o Direito Internacional: Uma Análise da Guerra e das Implicações do Uso da Inteligência Artificial

Autonomous Weapons Systems and International Law: An Analysis of War and the Implications of the Use of Artificial Intelligence

GILMAR ANTONIO BEDIN[1]
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí).

ALINE MICHELE PEDRON LEVES[2]
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí).

LAURA MALLMANN MARCHT[3]
Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí).

RESUMO: As transformações da guerra foram significativas com o fim do século XX e a emergência do século XXI. Este fato trouxe à tona a configuração de novas demandas para o Direito Internacional, sobretudo com o desenvolvimento de tecnologias bélicas revolucionárias. Neste sentido, o presente artigo reflete sobre o grande desafio que a sociedade e o Direito Internacional terão para garantir a manutenção da paz mundial e evitar as significativas consequências que a utilização dos sistemas de armas autônomas (AWS) poderia gerar. Por meio do método científico de procedimento hipotético-dedutivo, da técnica de pesquisa bibliográfica e da abordagem qualitativa, este trabalho exploratório objetiva analisar especificamente: i) os Estados soberanos e a guerra típica do mundo moderno; ii) as profundas transformações na sociedade globalizada e as novas formas ou meios de realização das guerras na ordem mundial atual; iii) os progressos tecnológicos de inteligência artificial (IA) do novo milênio com a criação dos sistemas de armas autônomas e a sua respectiva necessidade de regulamentação pelo Direito Internacional. Concluiu-se que este debate ainda é relativamente recente e que as normas internacionais atuais são insuficientes para a regulamentação da utilização dessas armas. Assim, é essencial a produção de um novo tratado internacional sobre o tema, tendo como parâmetro os princípios éticos e humanitários. Para tanto, precisa haver a produção de um

1 Orcid: http://orcid.org/0000-0001-9183-7065.
2 Orcid: http://orcid.org/0000-0002-0371-5234.
3 Orcid: http://orcid.org/0000-0002-0780-0452.

novo consenso entre os Estados e a superação das posições mais contundentes quanto à utilização desses sistemas de armas autônomas ou sua plena proibição, justificando-se o valor e a originalidade desta pesquisa.

PALAVRAS-CHAVE: Direitos humanos; Direito Internacional; guerras atuais; sistemas de armas autônomas; transformações políticas e tecnológicas.

ABSTRACT: The transformations of the war were significant with the end of the XX century and the emergence of the XXI century. This fact brought up the configuration of new demands for International Law, especially, with the development of revolutionary weapons technologies. In this sense, this article reflects about the great challenge that the society and International Law will have to guarantee the maintenance of world peace and avoid the significant consequences that the use of autonomous weapons systems (AWS) could generate. Through of the hypothetical-deductive procedure scientific method, the bibliographic research technique and the qualitative approach, this exploratory work aims to analyze specifically: i) the sovereign states and the typical war of the modern world; ii) the profound transformations in the globalized society and the new ways or means of wars in the current world order; iii) the technological advances in artificial intelligence (AI) in the new millennium with the creation of autonomous weapons systems and their respective necessity for regulation by International Law. It was concluded that this debate is still relatively recent and that current international rules are insufficient to regulation the use of these weapons. Thus, it is essential the produce of a new international treaty on the theme, having ethical and humanitarian principles as a parameter. Therefore, there needs to be the production of a new consensus among the States and the overcoming of the most forceful positions regarding the use of these autonomous weapons systems or their complete prohibition, justifying the value and originality of this research.

KEYWORDS: Human rights; International Law; current wars; autonomous weapon systems; political and technological transformations.

INTRODUÇÃO

A guerra e as armas utilizadas na sua realização acompanham os processos de transformação histórica das sociedades humanas. Partindo desse pressuposto, a história da humanidade pode ser contada por meio das guerras, as quais assumiram diversas formas, e, se por um lado, geraram um conjunto de consequências negativas – a destruição e a morte –, por outro, também contribuíram para o desenvolvimento de habilidades humanas – ciência e tecnologia. Do combate corpo a corpo e da cavalaria dos exércitos à criação dos instrumentos bélicos de alta precisão, do arco e das flechas até os canhões, mísseis, tanques de guerra, submarinos, aviões de combate, armamentos químicos, biológicos e nucleares do século XX, as

guerras atravessaram as civilizações e produziram consequências de proporções inimagináveis.

No início do século XXI, é possível perceber, a partir do desenvolvimento das novas tecnologias, uma nova e profunda transformação da guerra, de suas estratégias e de seus principais meios de realização. É que ela está deixando ser um conflito realizado essencialmente por grupos humanos militarmente treinados e passando a ter as feições enquanto uma forma de disputa entre computadores dotados de inteligência artificial (IA). Fato é que este consiste em um acontecimento extraordinário e que possui grandes e complexas implicações. Por isso, surge imediatamente a preocupação de diversos grupos presentes na sociedade internacional sobre como regular a utilização desses novos recursos bélicos e, em consequência, definir claramente a responsabilidade de todos os envolvidos. Sem dúvidas, trata-se de um dos grandes temas da sociedade atual e refletir acerca de seus contornos é a preocupação central do presente artigo.

As novas tecnologias mudaram, de fato, a lógica dos conflitos e a forma atual das guerras, uma vez que na contemporaneidade não é mais possível falar em guerras instrumentais típicas da sociedade internacional clássica ou moderna[4]. Daí, portanto, a importância dessa investigação científica, no sentido de que, com o avanço das novas tecnologias, o Direito Internacional precisa regulamentar a utilização dos sistemas de armas autônomas, pois têm enorme capacidade de destruição e de implicações que podem violar a dignidade humana e muitos dos direitos protegidos pela sociedade internacional. Em diferentes contextos, repensar as formas de guerra exige uma abordagem verdadeiramente interdisciplinar e, desse modo, o presente artigo problematiza: Em que medida a sociedade e o Direito Internacional conseguirão gerir os desafios para a manutenção da paz mundial e as imprevisibilidades de novos conflitos frente à utilização dos sistemas de armas autônomas como vias de solução das controvérsias?

A hipótese embrionária desta pesquisa reside no fato de que o progresso tecnocientífico impacta diretamente nas esferas ética e jurídica, de tal forma que se faz imprescindível a elaboração de um novo tratado para a sua regulamentação ou, então, a adaptação das normas de Direito Internacional existentes para que disciplinem a utilização ou a proibição da utiliza-

4 Não, pelo menos, no sentido clássico da guerra. Da guerra como a continuação da política por outros meios (Clausewitz, 2010).

ção dos sistemas de armas autônomas. Nota-se que o Direito Internacional Humanitário (DIH) ou Direito Internacional dos Conflitos Armados (DICA) consiste em um ramo do Direito Internacional Público, sendo instrumental e indissociável à temática e à problemática analisadas neste artigo científico. Esse consiste em um conjunto de normas, princípios gerais e costumes que regem as relações constituídas por acordos – tratados e convenções – entre os Estados, os quais impõem limitações aos possíveis efeitos dos conflitos armados, protegendo aqueles que não participam das hostilidades e restringindo os meios e/ou métodos de combate. Atualmente, pode-se afirmar que o DIH ou DICA é aceito como Direito Consuetudinário, isto é, como um conjunto de regras gerais que se aplicam a todos os Estados que compõem a sociedade internacional.

Metodologicamente, neste trabalho de caráter exploratório e de abordagem qualitativa, foi utilizado o método de procedimento hipotético-dedutivo e a técnica de pesquisa bibliográfica e documental indireta, a partir de uma revisão literária e análise crítica das referências utilizadas – livros, artigos científicos e *websites*. Para tanto, o artigo estrutura-se em três seções distintas que objetivam analisar especificamente: i) os Estados soberanos e a guerra típica da modernidade; ii) as transformações oriundas dos processos da globalização na sociedade contemporânea e as novas formas ou meios de realização das guerras; iii) os progressos tecnológicos do século XXI e a emergência dos sistemas de armas autônomas, as quais trazem à tona um conjunto de desafios em razão da potencialidade dessas máquinas e a necessidade de regulamentação pelo Direito Internacional.

1 OS ESTADOS SOBERANOS E A GUERRA MODERNA

A afirmação histórica dos Estados modernos está vinculada à ideia de soberania e ao monopólio da arte da guerra. Essa constatação fica evidente quando são retomados os acontecimentos que envolveram a chamada Guerra dos Trinta Anos[5] e a assinatura dos tratados da Paz de

5 Foi uma das mais longas e cruéis guerras centradas na Europa e que envolveu praticamente todos os Estados europeus em um conflito dinástico-religioso, travado por exércitos providos de poderes devastadores. Em princípio, a Guerra dos Trinta Anos localizou-se na parte central europeia, mais especificamente no território do Império alemão. Essa guerra, iniciada em 1618, foi a consequência direta do episódio em que o Sacro Império Romano-Germânico tentou aniquilar com os protestantes da Boêmia. Pouco a pouco, o conflito foi se generalizando em virtude de uma série de disputas de ordem econômica, territorial, política e religiosa. Com o término da guerra, em 1648, saíram vitoriosos: a França, a Suécia e os Estados protestantes; e perderam a guerra: a Espanha, o Sacro Império Romano-Germânico e a Santa Sé (Bedin, 2001; 2011).

Westfália[6]. Tais acontecimentos colocaram os Estados como o núcleo central da sociedade internacional e, em consequência, deixaram de reconhecer qualquer poder superior ao seu próprio. Além disso, os fatos referidos transformaram a possibilidade de realização da guerra em uma prerrogativa legítima dos Estados soberanos e em um ato completamente normal da sociedade internacional (Bedin, 2011).

Frente a isso, os Estados passaram a se constituir como verdadeiras potências e, a partir desse *locus* definido, criaram, propriamente, as primeiras normas de Direito Internacional Público[7]. Evidentemente, as referidas normas estavam voltadas para o reconhecimento do poder soberano dos Estados e para a afirmação da guerra como um de seus direitos. Nesse sentido, a sociedade internacional incorporou a experiência de normalização da guerra como uma forma regular de exercício do poder estatal, dos interesses nacionais e, inclusive, como um mecanismo de regulação social que, em última análise, poderia garantir a segurança e a paz no contexto das relações internacionais modernas[8]. Essa arquitetura se materializou na ideia de que o poder equilibra o poder como um sistema de balanças. Assim, surgiu, como bem lembra Norberto Bobbio (2009), um sistema que pode ser definido como um cenário típico de um "terceiro ausente" e que tinha como pressuposto fundamental a esperada mútua abstenção dos Estados, as quais se pautavam nos interesses nacionais dos países envolvidos e no sistema de equilíbrio de poder (Bedin, 2011; Lafer, 2012).

Em outras palavras, enquanto os Estados se afirmavam como unidades políticas centrais da sociedade internacional moderna, as disputas interestatais passaram a ser resolvidas pelo poder de cada país, fato esse que tornou tanto a violência quanto a força recursos legítimos do poder estatal. Por isso, esse momento histórico se configurou como uma espécie de anarquia, no qual todos os Estados detinham plena liberdade para fazer aquilo que julgavam ser mais favorável aos seus interesses. Em decorrência disso, Gilmar Antonio Bedin (2001, p. 352) afirma que a sociedade internacional do período possuía como único recurso disponível, "para o estabelecimento

6 A Paz de Westfália (1648), composta por um conjunto de tratados, foi o primeiro acordo internacional que afirmou a centralidade política do Estado moderno e estabeleceu os princípios norteadores da soberania estatal, da não intervenção e da separação entre as ordens políticas e religiosas. Portanto, os tratados westfalianos se configuram como os pilares de sedimentação do Direito Internacional clássico ou moderno enquanto um sistema jurídico baseado na soberania absoluta e indivisível dos Estados (Bedin, 2001; Menezes, 2005).

7 No sentido moderno desta área do conhecimento (Mello, 1997).

8 O raciocínio da afirmação da guerra como algo natural pode ter origem, também, no fato de que o próprio ser humano era considerado, ainda na modernidade, intrinsecamente propenso ao conflito (Schneewind, 2001).

de um cenário de paz duradoura, o sistema de equilíbrio ou balança de poder", por meio do qual o poder de cada Estado acabava por limitar a força dos demais, estabilizando, de certo modo, as relações internacionais "como um substituto de uma autoridade supranacional inexistente" (Moreira, 1996, p. 62). Portanto, na tentativa de estabelecer a paz, ainda que de forma negativa, esse sistema sempre reproduzia o *status quo* existente e, por conseguinte, novos conflitos por novas hegemonias.

Com isso, na sociedade internacional clássica ou moderna, de acordo com o entendimento de Jerome B. Schneewind (2001, p. 99), a guerra passou a ser compreendida como um fato completamente normal da vida civilizacional. Isso aconteceu porque, como o referido autor destaca, os seres humanos são "autopreservadores e briguentos". Consequentemente, a compreensão de que a guerra tem a violência como elemento originário adquiriu contornos objetivos, sendo caracteristicamente fruto do ódio e da animosidade. Desse modo, é possível dizer que Carl Von Clausewitz (2010, p. 30) tinha razão quando assegurou que a guerra é "como um cego impulso natural, depois, do jogo das probabilidades e do acaso, que fazem dela uma livre atividade da alma, e, finalmente, a sua natureza subordinada a instrumento da política". Foi no século XVIII que Clausewitz (2010) denunciou a fria realidade do conceito da guerra, no sentido de que esta é um fenômeno ambivalente e capaz de modificar a sua natureza de acordo com a circunstância concreta. Daí a sua afirmação de que a guerra consiste em um instrumento político, ou seja, na realização das relações políticas estatais por outros meios, a qual se traduz por meio do exercício da violência armada para eliminar a vida humana e qualquer forma de resistência. Isso posto, segundo o evidenciado por Michael Hardt e Antonio Negri (2005, p. 33), "a guerra vai-se transformando no princípio básico de organização da sociedade, reduzindo-se a política apenas a um de seus recursos ou manifestações". Essa é a natureza da guerra moderna.

Na realização da guerra moderna, os Estados soberanos se utilizavam de armamentos bastante desenvolvidos para a época, como o canhão e o mosquete – uma das primeiras armas de fogo portáteis. Mas foi com as Revoluções Americana pela Independência dos Estados Unidos da América (1776-1783) e Francesa (1789-1799) que os meios de guerra começaram verdadeiramente a se transformar, incorporando novas tecnologias nesse período. A partir de então, as táticas e estratégias das guerras foram largamente aprimoradas com sistemas inovadores, a exemplo do que ocor-

reu com as chamadas guerras napoleônicas (1803-1815)[9]. Em seguida, na Guerra da Crimeia (1853-1856) e, também, na Guerra Civil dos EUA (1861-1865), foram empregados rifles de maior alcance, o navio a vapor, além de novas táticas de movimentos e de trincheiras[10]; enquanto na Guerra Franco-Prussiana (1870-1871) os exércitos da Prússia passaram a se utilizar das linhas férreas de trens como logística para movimentar suas tropas militares (Ferroni, 2002).

Esses progressos foram extraordinários. Mas os verdadeiros e intensos avanços das tecnologias bélicas surgiram com a deflagração da Primeira Guerra Mundial (1914-1918) e a corrida armamentista dos Estados envolvidos – como os tanques blindados, os submarinos, as minas terrestres, as metralhadoras, as granadas explosivas, os gases tóxicos ou as armas químicas e os aviões de combate (Ferroni, 2002; Gannon, 2020). Os aviões, contudo, foram usados em pequena escala, pois, apesar de o primeiro grande conflito de ordem global ter envolvido tropas de infantaria, marinha e força aérea, a forma de guerra ainda era predominantemente terrestre (Gannon, 2020; Gilbert, 2017). Ao mesmo tempo, houve um vácuo tecnológico no tocante à substituição dos cavalos, isso porque qualquer alternativa a ela para a realização de movimentação nos campos de batalha, fosse para reconhecimento, recuo ou perseguição, era bastante limitada – a exemplo dos trens (que não perpassavam pelas zonas de conflito) e automóveis (que ainda eram uma promessa frágil). Além do mais, até mesmo os carros de combate ou tanques de guerra eram considerados recursos bélicos vulneráveis, lentos e deficitários no deslocamento e de difícil manutenção e reparo; já os aviões, como referido, não foram bem aproveitados estrategicamente para contribuir com as operações da guerra terrestre (Duarte, 2012; Phillips, 1999).

Mesmo que os avanços da Primeira Grande Guerra tenham sido, de fato, inovadores, foi com a Segunda Guerra Mundial (1939-1945) que se romperam todos os padrões de guerra até então existentes. Vale destacar que, apesar dos esforços internacionais para a limitação do Direito da Guerra, como as Conferências da Paz de Haia (1899 e 1907) e a Liga das

9 Fato esse que pode ser observado, por exemplo, nos escritos de Carl Von Clausewitz (2010).

10 As trincheiras consistiam em escavações lineares no solo, com aproximadamente dois metros de profundidade e centenas – ou até milhares – de quilômetros de extensão. Nesse sentido, a guerra de trincheiras foi uma forma terrestre de combate, na qual as tropas militares desfrutavam de uma posição considerada bem protegida nos campos de batalha contra-ataques de artilharias inimigas. Conforme afirmou Megan Gannon (2020, [s.p.]), "a guerra de trincheiras foi um componente crucial do cenário europeu na Primeira Guerra Mundial".

Nações (1919-1946), a sociedade internacional não conseguiu evitar a deflagração de outra guerra mundial. Assim, de uma forma completamente diferente do conflito anterior, os aviões foram aprimorados e passaram a ser utilizados em larga escala, fato esse que ocorreu, também, com os tanques e submarinos de guerra. Além disso, os exércitos envolvidos nos confrontos se utilizaram de radares, rádios transmissores, canhões, mísseis de longo alcance, foguetes, novos e mais potentes tipos de armas de fogo, bombas, minas e outros recursos explosivos. Com efeito, todas as transformações foram extraordinárias. No entanto, a tecnologia bélica mais inusitada e, simultaneamente, devastadora, foi empregada somente no fim da Segunda Guerra Mundial, em 1945, nas cidades japonesas de Hiroshima e Nagasaki. Essa nova tecnologia de guerra foi a bomba atômica[11] (Gilbert, 2014; Masson, 2017). A partir desse momento, o fenômeno da guerra ultrapassou a premissa da instrumentalidade política clausewitziana e introduziu a essência de uma nova e terrível possibilidade: a destruição em massa de civilizações inteiras mediante o vasto desenvolvimento das inovadoras tecnologias de energia nuclear, bem como das armas químicas e biológicas.

Desse modo, pode-se afirmar que as destrutivas tecnologias empregadas na maior guerra mundial provaram até que ponto os antagonismos e as rivalidades estatais westfalianas podem chegar quando vinculados aos nacionalismos exacerbados que se evidenciam em um panorama anárquico e repleto de desequilíbrios políticos, econômicos e sociais (Churchill, 2017; Tota, 2011). Apesar dos efeitos avassaladores da Segunda Guerra Mundial, esse mega conflito bélico foi, ao mesmo tempo, responsável pelo progresso e aprimoramento de uma série de inovações tecnológicas e científicas. A exemplo disso, tem-se os avanços da medicina e a obtenção da forma purificada da penicilina, importante antibiótico utilizado no tratamento de seres humanos; a ascensão da aviação para combater as forças inimigas com bombardeios aéreos de alto nível e a criação dos helicópteros pela indústria aeronáutica; a invenção do primeiro computador moderno, o qual foi empregado para realizar cálculos balísticos da marinha norte-americana; a produção de mísseis que, posteriormente, foram aperfeiçoados pela indústria espacial para levar o homem à lua (Gilbert, 2014; Masson, 2017).

11 A Segunda Guerra Mundial findou com o emprego da bomba atômica, a arma mais terrível já descoberta e resultante de processos políticos, tecnológicos e científicos. A partir de então, o alcance, a escala e os danos das bombas nucleares alteraram completamente o modo tradicional da guerra, do poder e das relações entre os Estados, provocando um intenso e constante medo por meio de uma ameaça que aflige a sociedade internacional com uma nova tecnologia de morte e destruição em massa (Gilbert, 2014).

Fato é que esse acontecimento bélico-político foi, com toda certeza, um marco histórico de vasta importância tecnocientífica, bem como para a construção do atual momento político internacional, mais interdependente e complexo. Destarte, com o término da maior guerra da história humana e com o amadurecimento das reflexões acerca das suas causas, tornou-se imprescindível a produção de um verdadeiro rearranjo das relações internacionais, com a gradativa relativização do conceito de soberania estatal e com a adoção de novas formas – pacíficas – para a solução dos conflitos interestatais. Foi este contexto que fortaleceu a ideia da construção de uma organização internacional de alcance mundial e o que impulsionou a criação da Organização das Nações Unidas (ONU) e seu Conselho de Segurança, em 1945.

Essa convergência pontual, contudo, não neutralizou uma tendência já presente no decorrer da Segunda Guerra Mundial: a crescente polarização entre duas novas superpotências emergentes, os Estados Unidos da América (EUA) e a União das Repúblicas Socialistas Soviéticas (URSS). Tal disputa criou um sistema internacional bipolar, o qual foi denominado de Guerra Fria (leste socialista *versus* oeste capitalista). Paradoxalmente, o resultado do pós-guerra estabeleceu as condições necessárias para a superação do cenário de "um terceiro ausente", mas, também, gerou novas incertezas e fortes tensões. Ao mesmo tempo, a sociedade internacional foi se transformando e novos atores internacionais foram se consolidando. Neste contexto, com a queda do Muro de Berlin, em 1989, e a dissolução da URSS, em 1991, uma nova ordem mundial multicêntrica e interdependente se revelou. Isso representou que o mundo dos Estados soberanos tradicionais estava relativizado e que o processo de globalização passava a ser uma nova realidade em concretização. Assim, para Anne-Marie Slaughter (2017), a visão de mundo pela política tradicional como um verdadeiro "tabuleiro de xadrez"[12], ou seja, um jogo interminável de poderes e vantagens estratégicos entre os Estados soberanos se alterou significativamente para uma visão de múltiplas conexões sociais, políticas, econômicas e culturais – entre nações, grupos,

12 "O tabuleiro de xadrez é uma metáfora tão dominante para ver e compreender o mundo dos Estados" (Slaughter, 2017, p. 6, tradução nossa). Texto original: "*The chessboard is such a dominant metaphor for seeing and understanding the world of States*", que Joseph Nye descreveu o mundo da política global pós-Guerra Fria como um "complexo jogo de xadrez tridimensional" (Nye, 2010, [s.p.], tradução nossa). Texto original: "*Complex three-dimensional chess game*".

redes e pessoas – dentro e fora das fronteiras territoriais e, inclusive, nas plataformas da *web*[13].

2 A SOCIEDADE GLOBALIZADA E AS NOVAS FORMAS DE GUERRA

Como ficou constatado na seção anterior, a chamada sociedade internacional clássica ou moderna estava superada. É que, de fato, o mundo não poderia continuar, diante de sua crescente complexidade e interdependência planetária, a se articular apenas a partir do poder soberano dos Estados e conviver com uma estrutura internacional propensa a gerar conflitos de grandes proporções. Desse modo, uma nova da ordem mundial se consolidou e passou a estar caracterizada por fluxos para além das fronteiras estatais e marcada pela participação de novos atores internacionais – Organizações Internacionais (OI), Organizações Não-Governamentais (ONGs) e Empresas Transnacionais (ETNs). Isso revelou, também, que a ordem estatal tradicional moderna havia sido progressivamente fragilizada e se adaptado a um novo cenário que, segundo Yuval Noah Harari (2018, p. 215), "nenhum Estado soberano é capaz de superar sozinho". Essa transformação aponta que a soberania estatal foi, aos poucos, deixando de ser absoluta e, evidentemente, nas palavras de Anne-Marie Slaughter (2017, p. 224), que "as bases da própria soberania estão mudando"[14] e se ajustando a uma realidade em que os Estados competem por poder e influência com muitos outros atores internacionais.

Com o fim do século XX, essas alterações globais se tornaram ainda mais intensas. Esse novo passo se alicerçou na solidificação das inovações das novas tecnologias e em sua capacidade extraordinária de disrupção. Além disso, a convergência de tais fatores produziu uma grande transformação nas formas de guerra, as quais assumiram dimensões globais, mesmo que estivessem restritas a determinadas localidades, visto que seus efeitos poderiam se alastrar mundialmente. Nesse sentido, de acordo com Hardt e Negri (2005, p. 21-22), a guerra, de fato, se transformou "num fenômeno geral, global e interminável", ao passo em que "inúmeros conflitos armados se manifestam hoje através do planeta, alguns breves e limitados a um lu-

13 No mundo da *web*, descrito por Anne-Marie Slaughter (2017), não há separação ou demarcações geográficas de fronteiras dos Estados soberanos, mas há inúmeras conexões, densidades e intensidades dos laços através das fronteiras. Para a autora, visualizar o sistema internacional como uma teia consiste em ver um mundo não de Estados, mas de redes que se cruzam e se sobrepõem em alguns lugares e são mais dispersas em outros.

14 Tradução nossa. Texto original: "*The foundations of sovereignty itself are shifting*".

gar específico, outros prolongados e expansivos". Então, na atual sociedade de risco, expressão cunhada pelo sociólogo Ulrich Beck (2011), sucessivos conflitos armados ocorrem nas mais diversas ordens e conjunturas sociais, os quais se aproveitam das avançadas criações tecnológicas da indústria bélica para alavancar suas magnitudes de destruição, como também se concretizar em qualquer lugar e a qualquer momento.

Diferentemente da era moderna, que se alicerçava nos ideais de ordem e estabilidade, a sociedade globalizada é caracterizada por incertezas e mutações constantes. Neste contexto, é expressivo que o incremento contínuo das tecnologias produz impactos que repercutem em todos os aspectos da vida, uma vez que essas podem ser caracterizadas por um grau relativamente elevado de riscos e contingências que lhes são próprios. Além do mais, a articulação das novas tecnologias, desenvolvidas rapidamente, envolve uma densa rede de movimentação de bens, de capitais e de fluxos pessoais e informacionais. Isso significa que os progressos tecnológicos e científicos[15] resultantes dos processos da globalização estão moldando a atual visão de mundo, assim como a compreensão e o comportamento dos indivíduos. Logo, novas fronteiras passam a integrar a vida em sociedade, na medida em que ocorre uma verdadeira desterritorialização instantânea do globo (Ianni, 2014).

Em termos de percepção, pode-se afirmar que a complexidade na qual a atualidade está arraigada é imensurável frente às constantes metamorfoses[16] mundiais e à exposição considerável dos riscos em um panorama integrado e transfronteiriço (Beck, 2018). Logo, é evidente que as tecnologias passam a se situar no epicentro das relações políticas, econômicas, culturais e sociais, o que influencia na conformação de uma nova ordem do globo, a qual reformula a ótica dos sistemas de estratégias para enfrentar as ameaças da sociedade internacional. O certo é que, segundo Pierre Lévy (2010, p. 16), "a tecnociência produziu tanto o fogo nuclear como as redes interativas", e, justamente por isso, a análise dos avanços das novas tecnologias não consiste apenas em avaliar os seus impactos, mas de "situar as

15 Conforme o entendimento de Andrew Feenberg (2017), a tecnologia e a ciência se tornaram dominantes por meio de disciplinas não técnicas, como o discurso dominante.

16 A metamorfose do mundo consiste em uma teoria plenamente original e que se destina a analisar por que a sociedade está cada vez mais difícil de ser compreendida. "A metamorfose, em contraposição, desestabiliza as certezas da sociedade moderna. Ela desloca o foco para 'estar no mundo' e 'ver o mundo', para eventos e processos não intencionais, [...] que prevalecem além dos domínios da política e da democracia [...] Metamorfose, nesse sentido, significa simplesmente que o que foi impensável ontem é real e possível hoje" (Beck, 2018, p. 11-12).

irreversibilidades às quais seus usos levariam" (Lévy, 2010, p. 26), tais como os riscos de utilização das invenções que oscilam à beira das catástrofes das guerras e ameaçam solapar, em parte, o Direito Internacional e os Direitos Humanos.

Dito isso, com os sucessivos e extraordinários progressos modernizadores das renovadas e complexas tecnologias, surgiram sistemas de armas completamente inéditos, os quais transfiguraram drasticamente as formas da guerra contemporânea. Neste ínterim, destaca-se que, já na década de 1990, durante a chamada Guerra do Golfo foi possível observar o uso de armas de altíssima tecnologia e uma celeridade incomparável na movimentação de tropas tradicionais (Zolo, 2009). Isso ficou ainda mais evidente com os ataques realizados pelos Estados Unidos, no Afeganistão em razão dos atentados terroristas de 11 de setembro de 2001[17]. A partir desse momento, já era possível observar que os investimentos em tecnologias bélicas aumentariam cada vez mais, pois os inimigos das chamadas guerras globais passaram a atuar de qualquer lugar do mundo. De fato, as novas tecnologias revolucionaram e ampliaram os métodos de guerra – como no caso da automatização e da robótica – e, consequentemente, lançaram produções inovadoras que envolveram os ramos da microeletrônica, da informática, da ciberciência, da física nuclear, da química fina e da biotecnologia nos atuais conflitos do século XXI.

Assim, essas amplas transformações reorganizaram as táticas e estratégias dos exércitos militares, produzindo efeitos expressivos que remodelaram os confrontos bélicos armados. Sob esse viés, Baldassarri e Nascimento (2020, p. 190-1) investigaram que as guerras, ao longo da história civilizacional, podem ser classificadas fundamentalmente em cinco tipos: 1. As guerras nas quais a massa humana dos exércitos estatais consiste no fator determinante e decisivo do poder de combate, assim como da vitória; 2. As guerras que relativizaram o conceito de massa e, por conseguinte, valorizaram o poder de fogo com a invenção do fuzil, das metralhadoras, do carregamento pela culatra e das peças de artilharia com tiros indiretos;

17 O início do novo milênio foi marcado pela "ocorrência do maior e mais devastador atentado terrorista da história, que mergulhou o mundo em um estado de medo generalizado, acarretou a morte de milhares de pessoas e destruiu o mais imponente símbolo do capitalismo global" (Leves; Bedin, 2019, p. 10). Em síntese, foi realizada, nos Estados Unidos da América (EUA), uma série de atentados terroristas arquitetados pela organização internacional fundamentalista islâmica *Al-Qaeda* (A Base) e sob a liderança de Osama Bin Laden, em reação à instalação de bases militares norte-americanas na península arábica durante a Guerra do Golfo (Leves; Bedin, 2019).

3. As guerras que associaram, de forma inigualável, a massa e o poder de fogo nos confrontos; contudo, a mecanização passou a ser o fator decisivo nos campos de batalhas não lineares, com o uso intensivo de inovações como os carros blindados de combate, os aviões e os submarinos controlados por sistemas de radiofrequência; 4. As guerras que não consideram os aspectos anteriores como essenciais, isso porque a informação passa a consistir no fator determinante ao utilizar computadores, Internet, pulsos eletromagnéticos, *lasers* e robôs em um contexto cibernético espacial; 5. As guerras que são, hoje, intensificadas com o intuito de influenciar a opinião pública mediante, por exemplo, as redes sociais e nas quais se fazem presentes as tecnologias integradas (como os drones), as biotecnologias e as nanotecnologias.

Evidentemente, as tecnologias de IA evoluíram muito ao longo das últimas duas décadas. Em função disso, os setores da indústria bélica de produção de armas e recursos militares não deixaram de acompanhar tal evolução. Inicialmente, apenas as munições se utilizavam de fontes de calor, raios *laser* e sinais de GPS[18] para a condução das armas, mas, "com o passar do tempo, a tecnologia [de guerra] evoluiu chegando ao ponto de haver um sistema inteiro robotizado e programado para atingir os alvos inimigos" (Baldassarri; Nascimento, 2020, p. 192). Esse método passou a ser denominado de sistemas de armas autônomas – *autonomous weapons systems* (AWS – sigla em inglês) e, segundo Guglielmo Tamburrini (2016, p. 125), "qualquer sistema de armas é autônomo, quando, uma vez ativado, pode selecionar e engajar alvos sem a intervenção adicional de um operador humano"[19]. Com efeito, as forças armadas dos Estados começaram a se interessar pelos sistemas de armas autônomas (AWS) não tripulados, os quais passaram a substituir as massas de exércitos humanos, uma vez que, conforme Christof Heyns (2016, p. 4), "assumiram a forma de drones armados e outros dispositivos de controle remoto, o que permitiu que os seres humanos estivessem fisicamente ausentes do campo de batalha"[20]. Neste contexto, tais sistemas possuem a capacidade de aderir a um objetivo sem

18 Sigla em inglês para o *Global Positioning System* (Sistema de Posicionamento Global), que compõe aparelhos receptores móveis, os quais utilizam os sinais de navegação por satélite.

19 Tradução nossa. Texto original: "*Any weapons system is autonomous when, once activated, can select and engage targets without further intervention by a human operator*".

20 Tradução nossa. Texto original: "*Took the form of armed drones and other remote-controlled devices, which allowed human beings to be physically absent from the battlefield*".

que haja desgaste político de forças humanas amigas, sendo, sob essa ótica, menos custosos quando comparados aos sistemas tripulados.

Até o presente momento, as decisões de empregar a força em um conflito armado ainda são tomadas pelos agentes humanos, embora essa ação possa ser executada à distância. Contudo, nota-se que, na atualidade, o significativo aumento da autonomia dos sistemas de armas aponta para a possível emergência de uma nova era, da chamada guerra 2.0, na qual os "humanos serão capazes de não apenas estar fisicamente ausentes do campo de batalha, mas também psicologicamente, no sentido de que os computadores determinarão quando e contra quem a força será liberada"[21] (Heyns, 2016, p. 4). Nesse quadro, há uma despersonalização do uso da força que, até então, era liberada por meio de controles remotos, sendo essa, portanto, elevada a um outro nível com a introdução de máquinas inteligentes dotadas de uma capacidade autônoma de selecionar alvos inimigos de forma independente e de lançar armas sem qualquer intervenção humana adicional. Até bem pouco tempo, essas tecnologias bélicas com visão computacional e IA existiam apenas no âmbito da ficção. Hoje, esses avanços se tornaram uma realidade com a utilização de drones de combate, soldados-robôs programados para lutar em zonas de conflito, óculos de realidade virtual capazes de detectar o inimigo, mísseis e armas com controle de tiros que disparam autonomamente quando selecionam alvos[22], veículos de combate terrestre que possuem a capacidade de identificar e engajar alvos muito mais rapidamente que os processos manuais, entre outros (Trindade, 2019).

É possível observar que algumas das principais potências mundiais – EUA, China, Rússia e o Reino Unido – estão investindo muito em IA e no aprendizado de máquinas com o objetivo de desenvolver sistemas aptos para selecionar os alvos e atirar. Sem embargo, a completa autonomia de máquina ainda não foi empregada em conflitos contra seres humanos, haja vista que essas ainda são controladas remotamente, mas já se alcançou o ponto em que tal possibilidade se tornou factível (Heyns; Sharkey, 2016). À vista disso, se faz imprescindível o questionamento acerca de como a so-

21 Tradução nossa. Texto original: "*Humans will be able to be not only physically absent from the battlefield but also psychologically absent, in the sense that computers will determine when and against whom force is released*".

22 Os sistemas de armas autônomas possuem sensores que fornecem um determinado grau de consciência situacional, nos quais os computadores processam as informações e, por sua vez, os efetores (armas) implementam as decisões computacionais com humanos fora do circuito (Heyns, 2016; Lieblich; Benvenisti, 2016).

ciedade e o Direito Internacional irão, respectivamente, responder e regular esse novo cenário composto por sistemas de armas autônomas letais, as quais poderão, muito em breve, ter o poder de acarretar graves ferimentos ou até mesmo a morte de seres humanos. Essa questão traz à tona implicações de ordens éticas, morais e jurídicas, bem como um conjunto de preocupações técnicas e operacionais, uma vez que existem diversos motivos pelos quais os sistemas de armas autônomas estão sendo desenvolvidos. O principal argumento seria a capacidade de proteção desses sistemas, pois eles "podem ser mais rápidos no engajamento dos alvos pretendidos porque conseguem processar informações mais rapidamente. Além disso, às vezes também é argumentado que os poderes superiores de processamento dos AWS podem evitar que alvos errados sejam atingidos"[23] (Heyns, 2016, p. 6-7). Na medida em que essas argumentações estão corretas, a utilização de armas autônomas poderia salvar vidas ou prevenir lesões injustificadas[24].

Igualmente a qualquer outra tecnologia, a IA apresenta tanto usos bons quanto ruins. No âmbito do policiamento e das forças armadas, a automação já é bastante utilizada para fazer trabalhos de limpezas e, também, no desmonte de bombas, ou seja, os robôs são aproveitados para fazer trabalhos maçantes, perigosos e sujos (Heyns, 2016). Mas, em um futuro próximo, os exércitos militares e os Estados desenvolvedores dos referidos AWS pretendem utilizá-los para evitar os erros e as falhas humanas, isso porque os soldados são, supostamente, menos confiáveis e precisos, bem como estão sujeitos ao cansaço e à problemas de ordem psicológica (Trindade, 2019). Nesse sentido, as máquinas efetuam um processamento de raciocínio automático, o qual se utiliza de respostas rápidas obtidas em treinamentos de repetição que estão voltados para a execução de determinadas tarefas. De outra perspectiva, o processamento controlado por seres humanos é mais lento e se refere ao raciocínio deliberativo, necessário na tomada de decisões cuidadosas em circunstâncias imprevisíveis e que se alteram dinamicamente. Em tais situações, o raciocínio automático dos sistemas de armas autônomas é debilitado e, portanto, o processo de seleção e engajamento

23 Tradução nossa. Texto original: "*May be quicker at engaging intended targets because they can process information faster. Furthermore, it is also sometimes argued that the superior processing powers of AWS can prevent the wrong targets from being hit*".

24 Um dos exemplos trazidos por Christof Heyns (2016, p. 7) reside nas situações de conflitos, quando os seres humanos podem, por medo, se antecipar no ataque e, nesse processo, ocasionar a morte de civis que não estão envolvidos nas hostilidades armadas.

de alvos letais pode dar errado em razão do mal funcionamento quando há informações contraditórias, isso, pois, a dúvida e a incerteza serão sempre suprimidas pela reação das máquinas, as quais costumam agir em questão de milissegundos (Sharkey, 2016).

A automação das armas pode, indiscutivelmente, possibilitar uma maior velocidade e precisão na seleção dos alvos ou, ainda, prevenir o uso excessivo da força. Contudo, as limitações das tecnologias de IA são preocupantes, visto que o erro pode ocasionar graves problemas humanitários e inflamar os conflitos. Para Noel Sharkey (2016, p. 34), "o ponto aqui é que é de vital importância que o raciocínio deliberativo humano seja habilitado no projeto de controle de supervisão para sistemas de armas. Embora esse também esteja sujeito a erros e falhas"[25]. Mesmo que alguns países, como os Estados Unidos e o Reino Unido, tenham estabelecido que sempre haverá um ser humano envolvido nos processos de decisões de letalidade, isto é, que os sistemas de armas autônomas e semiautônomas estarão sob controle humano, sendo projetados para que seus operadores exerçam níveis adequados de deliberação e julgamento para o uso da força, o que não restou evidente é o tipo de controle que será empregado. Desse modo, não houve transparência na definição das operações das armas computadorizadas, pois a simples afirmação de que um ser humano está presente no circuito de controle não esclarece, devidamente, o seu grau de envolvimento. De um lado, isso pode significar que o humano está simplesmente programando o sistema de armas para determinada missão ou pressionando um botão para ativá-lo; de outro, de maneira um tanto quanto esperançosa, pode significar que o ser humano está exercendo um julgamento interpretativo completo acerca da legitimidade de um alvo antes de dar início ao ataque (Sharkey, 2016). Por essas razões, a IA vem adquirindo uma atenção crescente na atual sociedade internacional globalizada em virtude das novas formas ou dos meios de realização das guerras.

3 OS SISTEMAS DE ARMAS AUTÔNOMAS E O DIREITO INTERNACIONAL

A atenção referida anteriormente vem acompanhada por um conjunto diverso de questões. Essas questões possuem algumas implicações técnicas, mas as mais relevantes têm uma dimensão ética, política e jurídica. Por se

25 Tradução nossa. Texto original: "*The point here is that it is vitally important that deliberative reasoning is enabled in the design of supervisory control for weapons systems. Although this is also subject to error and flaws*".

tratar de uma tecnologia abrangente, a inteligência artificial tem o potencial de afetar múltiplos aspectos da vida humana, pois está impulsionando a ampliação da utilização de sistemas de armas autônomas letais (LAWS) nos conflitos armados (Wischmeyer, 2020). É que, como lembra Ugo Pagallo (2013, p. 58, tradução nossa), "a capacidade dos robôs de operar no mundo real sem controle humano tem impacto sobre o princípio fundamental das leis da guerra", uma vez que "tanto a autonomia quanto a imprevisibilidade do comportamento das máquinas de IA tornam as guerras de robôs profunda e irremediavelmente antiéticas". Desse modo, o autor ressalta que, "se os robôs causariam sérios danos ao tomar suas próprias decisões, esse seria apenas um pequeno passo para imaginar robôs que podem provocar guerras acidentais" (Pagallo, 2013, p. 58).

Há, portanto, grandes implicações no uso dos sistemas de armas autônomas nos conflitos entre os seres humanos. Isso se dá porque, como já foi referido, os direitos à vida e à integridade física das pessoas podem ser violados de forma aleatórias, especialmente nas situações que os computadores não são capazes de identificar corretamente os alvos ou não conseguem distinguir civis e combatentes ou, ainda, identificar aqueles que estão feridos e/ou em processo de rendição. Além disso, nota-se que, mesmo nas situações mais adversas, os seres humanos são capazes de agir segundo os princípios da proporcionalidade e precaução – no sentido de que os eventuais danos colaterais infligidos em civis não devem ser excessivos em relação às vantagens militares obtidas (Heyns, 2016). Isso vai depender sempre do sentido do humano e de um sistema ético de moralidade.

Os sistemas de armas autônomas, por mais desenvolvidos que possam ser, nunca terão esses recursos e, portanto, consistem em uma verdadeira ameaça à dignidade humana. Nas palavras de Heyns (2016, p. 11), uma máquina, com sistema de arma autônoma, "sem sangue e sem moralidade ou [sentido de] mortalidade não pode fazer justiça à gravidade da decisão de usar a força em um caso particular, mesmo que seja mais precisa do que os humanos"[26]. Essa limitação ocorre porque a "morte por algoritmo significa que as pessoas são tratadas simplesmente como alvos e não como seres humanos completos e únicos"[27] (Heyns, 2016, p. 11). De certo modo, há uma

26 Tradução nossa. Texto original: "*Which is bloodless and without morality or mortality, cannot do justice to the gravity of the decision whether to use force in a particular case, even if it may be more accurate than humans*".

27 Tradução nossa. Texto original: "*Death by algorithm means that people are treated simply as targets and not as complete and unique human beings*".

exigência implícita nas normas e nos princípios éticos do Direito Internacional de que apenas os humanos são capazes de tomar a decisão de empregar força contra outros da mesma espécie, haja vista que os computadores, os robôs e as demais máquinas não podem ser programados e tampouco são aptos para responder adequadamente às infinitas possibilidades e cenários que a vida humana oferece. Porém, os fóruns internacionais, nos quais essas novas tecnologias de armas autônomas com IA deveriam ser discutidas, são estrutural e substancialmente incapazes de lidar com os progressos exponenciais das aplicações bélico-militares (Trindade, 2019). Assim, enquanto uma regulamentação efetiva de AWS entre os Estados não surge no âmbito do Direito Internacional, as incertezas das guerras permanecem.

Diante deste cenário, alguns grupos começaram a se posicionar contra o desenvolvimento dos sistemas de armas totalmente autônomas e letais. Entre os principais, está o manifesto denominado "*Stop Killer Robots*" ("Parem os Robôs Assassinos"), campanha promovida pela *Human Rights Watch*, Organização Não Governamental (ONG) de Direitos Humanos. Tal organização aponta que, atualmente, seis países – EUA, China, Israel, Coréia do Sul, Rússia e Reino Unido – estão avançando rapidamente no desenvolvimento de armas com autonomia significativa ou total nas funções de seleção e ataque de alvos, capazes de liberar a força sem qualquer controle humano por detrás do sistema. Assim, conforme consta na página *on-line* (*website*) da *Stop Killer Robots* (2021, [s.p.]), "se [isso] não for controlado, o mundo poderá entrar em uma corrida armamentista robótica desestabilizadora"[28]. Essa campanha cita a Declaração e Programa de Ação de Viena, publicada na Conferência Mundial sobre Direitos Humanos de 1993, pois, conforme esse documento internacional, "todos os direitos humanos têm origem na dignidade e valor inerente à pessoa humana" (OEA, 1993, p. 2) e, por conseguinte, de acordo com Dieter Birnbacher (2016, p. 107), as "armas totalmente autônomas podem minar o princípio da dignidade, o que implica que todos têm um valor que merece respeito"[29], até porque as máquinas não possuem características sentimentais – tais como a compaixão – necessárias para fazer escolhas éticas consideradas complexas. Além do mais, "uma forte razão contra a transferência do conceito de dignidade humana para objetos abstratos é a implausibilidade de considerar

28 Tradução nossa. Texto original: "*If left unchecked the world could enter a destabilizing robotic arms race*".

29 Tradução nossa. Texto original: "*Fully autonomous weapons could undermine the principle of dignity, which implies that everyone has a worth deserving of respect*".

as violações dessa dignidade hipotética com a mesma gravidade que as violações da dignidade individual"[30] (Birnbacher, 2016, p. 110).

Apesar disso, os debates acerca da proibição ou da regulamentação restritiva do uso das armas autônomas pelo Direito Internacional avançaram muito pouco e, inclusive, só costumam receber alguma atenção em razão da pressão realizada pelos movimentos ativistas. Isso ocorre, especialmente, porque aqueles que fabricam e investem nesse tipo de tecnologia bélica com IA não estão muito interessados nos diálogos que podem prejudicar consideravelmente os seus interesses. Segundo Guilherme de Aguiar Patriota[31], em resposta aos questionamentos feitos por Rodrigo Trindade (2019, [s.p.]), na reportagem especial do canal Tilt – sobre tecnologia do UOL –, "poucas armas foram proibidas até hoje, porque armamentos têm complexos industriais, tecnológicos e interesses por trás", desse modo, tais setores "não escondem de ninguém que não apoiam um instrumento internacional juridicamente vinculante que proíba ou restrinja o desenvolvimento dessas armas [autônomas]". É justamente por isso que esse tema, desde 2013, ainda está em discussão no Conselho de Direitos Humanos da ONU e não obteve sucesso nas deliberações acerca da violação das normas fundamentais do Direito Internacional Humanitário (DIH) pelos AWS (Trindade, 2019). Assim, se, por um lado, a sociedade internacional segue a passos lentos, com poucos Estados dispostos a buscar novos limites de regulação dos sistemas de armas autônomas; de outro, aqueles que investem na fabricação desse tipo de tecnologia de IA, que amplia o poderio militar, avançam rapidamente.

Ademais, todos que defendem o uso das armas autônomas afirmam que essa tecnologia possui uma espécie de autorregulamentação própria, fato esse que faria com que esses sistemas respeitassem o DIH de uma certa forma até melhor do que os próprios combatentes humanos. Entretanto, este consiste em um argumento retórico das grandes potências militares e que não se sustenta, uma vez que já está comprovado que as atuais tecnologias com reconhecimento de imagens, fundamentais na utilização dos sistemas de armas autônomas, podem ser imprecisas e apresentar, inclusive, tendências racistas, xenófobas e excludentes. Nesse sentido, as argumen-

30 Tradução nossa. Texto original: "*A strong reason against the transference of the concept of human dignity to abstract objects is the implausibility of regarding violations of this hypothetical dignity with the same gravity as violations of individual dignity*".

31 Foi embaixador representante especial do Brasil junto à Conferência de Desarmamento em Genebra e presidente do grupo de peritos sobre segurança cibernética da Organização das Nações Unidas (ONU).

tações éticas voltadas para banir os AWS demonstram que não importa o quão bem esses sistemas venham a realizar suas tarefas de direcionamento e engajamento ou se estão em conformidade com as normas do Direito Internacional e dos Direitos Humanos, pois "existem razões morais imperiosas para proibir o uso"[32] (Tamburrini, 2016, p. 137). Ainda, de acordo com Birnbacher (2016, p. 107), outro recurso que pode ser levado ao extremo pelos sistemas de armas totalmente autônomas e letais consiste no risco da imprevisibilidade dos ataques que exacerbam a ameaça dessas máquinas para os humanos, isso em razão de que "eles próprios decidem quem será o objeto de ataque, usando sua inteligência embutida e sem qualquer controle direto de uma unidade de controle humano"[33].

Frente a isso, não há qualquer garantia de que o uso da força não seja arbitrário e imprevisível sem o controle humano, sua supervisão e responsabilidade, sendo, portanto, considerado imoral matar sem o envolvimento da razão humana, pois isso reside em um ato unilateral e completamente injustificado. Esta consiste, então, em "uma posição deontológica sofisticada, baseada em considerações emanadas do DIH e complementadas pela ética"[34] (Lieblich; Benvenisti, 2016). Para a preservação da moralidade, da dignidade, da justiça e da lei, consoante Tamburrini (2016, p. 138), "não podemos aceitar um sistema automatizado tomando a decisão de tirar uma vida humana"[35]. Em particular, os debates jurídicos sobre a possibilidade de os AWS serem utilizados em conformidade com as regras de conduta de hostilidades elencadas pelo Direito Internacional Humanitário estão, atualmente, amparados em uma previsibilidade futura sobre a sofisticação tecnológica das armas, suas capacidades e circunstâncias de uso, bem como acerca das projeções de controle. Todavia, diante da falta de transparência – que é bastante limitada – por parte dos Estados que investem e desenvolvem essas tecnologias, a sociedade internacional não tem como saber, precisamente, se os sistemas de armas autônomas são – ou serão – usados

32 Tradução nossa. Texto original: "*There are overriding moral reasons to forbid their use*".

33 Tradução nossa. Texto original: "*They decide themselves who is to be made the object of attack, by using their inbuilt intelligence and without any direct control by a human control unit*".

34 Tradução nossa. Texto original: "*A sophisticated deontological position, based on considerations emanating from IHL and supplemented from ethics*".

35 Tradução nossa. Texto original: "*Cannot accept an automated system making the decision to take a human life*".

em conformidade com os requisitos do Direito Internacional dos Conflitos Armados[36] (DICA).

Conforme evidencia Sarah Knuckey (2016), os proibicionistas dos referidos sistemas argumentam que as armas autônomas não seriam capazes de cumprir o DIH e, consequentemente, defendem a criação de um novo tratado internacional que proíba expressamente os AWS. Já, na outra extremidade, os apoiadores desses sistemas argumentam contra a proibição e apresentam possíveis circunstâncias ou políticas que estruturariam a legalidade do uso das tecnologias de armas autônomas. Essa discussão, no entanto, é muito mais ampla do que ambas as linhas argumentativas, pois, no cerne das objeções pessimistas encontra-se um profundo ceticismo em relação à capacidade de os computadores atingirem um nível de sofisticação necessário para seguir as complexas normas do DIH; já os otimistas são enfáticos ao afirmar que a humanidade será melhor com a utilização desses sistemas de armas autônomas, visto que eles são imunes aos aspectos negativos da psique humana – tais como a vingança, o preconceito e o medo –, que podem resultar em riscos materiais para civis. Tendo em vista que o Direito Internacional Humanitário exige dos beligerantes um cuidado constante para com a vida dos civis, os pessimistas são rápidos para replicar que a eliminação do fator humano pode, sim, neutralizar as características adversas, mas, também, faz o mesmo com as positivas, as quais podem permitir com que os seres humanos ajam de forma deliberativa e com misericórdia (Lieblich; Benvenisti, 2016). Isso é típico dos seres humanos. Assim, por mais que alguns processos mentais, em seus aspectos epistemológicos, possam ser reproduzidos por *softwares*, isso não quer dizer que a inteligência, característica própria do ser humano, possa ser completamente substituída por máquinas, uma vez que as decisões humanas são tomadas a partir da estrutura psíquica e biológica dos indivíduos, que está imbuída de experiências socioculturais (Pérez Luño, 1996). Os riscos são imensos e, portanto, é necessária uma nova regulamentação sobre o tema.

36 O Direito Internacional dos Conflitos Armados ou Direito Internacional Humanitário consiste em um conjunto de normas internacionais, de ordem convencional ou consuetudinária, destinadas especificamente a serem aplicadas nas situações de conflitos armados. Essas normas limitam, por razões humanitárias, o direito de as partes envolvidas em conflito escolherem, livremente, os métodos e os meios a serem empregados na guerra. Entre os requisitos para a utilização de armas pelo DICA ou DIH – incluindo aqui nesta análise as autônomas, mesmo que não expressas em lei – estão: os princípios da proporcionalidade e discriminação; o seu emprego de forma restrita ou integrada ao projeto do sistema; as proteções contra possíveis falhas – devendo no caso das armas autônomas haver uma intervenção humana antes que níveis inaceitáveis de danos ocorram; a capacidade de cumprir, apenas, a missão militar (Corn, 2016).

Outrossim, é uma tarefa urgente e imprescindível a celebração de um novo tratado internacional que regulamente o uso de armas autônomas ou, simplesmente, que torne esses instrumentos proibidos. No caso da regulamentação, é fundamental que a sua utilização esteja sobre um controle humano[37] significativo no circuito de operações e que permita a consideração de questões como a dignidade humana (Bhuta; Beck; Geiß, 2016). Essa iniciativa tem o apoio dos países que integram a União Europeia (UE), que tem atuado em conjunto com a Organização das Nações Unidas (ONU) para torná-lo, de fato, o princípio regulador deste tema (Roff; Moyes, 2016). De qualquer forma, o importante, neste contexto, é chamar a atenção para a urgência e para a existência de um grande problema: a lacuna existente na ordem jurídica internacional acerca da responsabilização legal – civil ou criminal – dos Estados em relação às consequências de suas ações. No direito civil, a principal questão consiste em como lidar com os eventuais danos causados pelos AWS; por sua vez, o direito penal está baseado na conduta – culposa ou dolosa – do ofensor da norma, em sua intenção de negligência ou imprudência sobre a violação de pessoas ou bens (Bhuta; Beck; Geiß, 2016; Birnbacher, 2016). Obviamente, não é possível atribuir responsabilidade pela conduta e suas consequências às armas autônomas. No entanto, essas armas não são completamente independentes da autoria humana, ou seja, sempre existem pessoas por detrás delas, seja no seu desenvolvimento ou na decisão de usá-las. Para Nehal Bhuta, Susanne Beck e Robin Geiß (2016, p. 357), "sem humanos agindo e tomando decisões, a cadeia de causalidade desde o projeto à implantação e o uso da força não pode ser concluída"[38]. Em síntese, são os indivíduos que respondem moral e legalmente pelo uso e pela ativação dos sistemas de armas autônomas, bem como pelos efeitos gerados a partir das operações.

Frente a isso, devem existir normas internacionais claras acerca do emprego das tecnologias de AWS e sobre a responsabilidade de todos os envolvidos. Isso, infelizmente, ainda não existe e a amplitude de sua complexidade exige a rápida celebração de um tratado internacional detalhado e abrangente. De fato, a complexidade desses sistemas dotados de inteli-

37 A discussão sobre o controle humano no contexto dos sistemas de armas totalmente autônomas é um tanto quanto paradoxal, uma vez que, se existe um controle humano por detrás das máquinas, isso significa que não pode haver uma autonomia total (Bhuta; Beck; Geiß, 2016). Consequentemente, a autonomia total existe quando os humanos não mais exercem um controle significativo (Heyns, 2016).

38 Tradução nossa. Texto original: "*Without humans acting and making decisions, the chain of causation from design to deployment to use of force could not be completed*".

gência artificial (IA) envolve diversos componentes humanos e tecnológicos (uma verdadeira interface homem-máquina) e pressupõe a colaboração de muitas partes envolvidas. Por isso, a responsabilidade deve ser precisamente fixada, pois será um enorme desafio definir o que ou quem está errado quando algo der errado (Bhuta; Beck; Geiß, 2016). Notadamente, essa lacuna de responsabilidade torna mais difícil garantir justiça, especialmente para as eventuais vítimas dessas armas autônomas e, além disso, a falta de responsabilização constitui uma verdadeira violação do DIH, sobretudo dos direitos à vida e à dignidade humana. Nessa mesma linha de reflexão, segundo Heyns (2016, p. 12), a responsabilização jurídica pode assumir diversas formas, "incluindo processos criminais, danos civis, medidas disciplinares ou a oferta de reparação ou compensação. Além da responsabilidade individual, instituições, como Estados ou corporações privadas, podem ser responsabilizadas"[39].

Com efeito, as alterações legais e as suas consequências apenas serão compreendidas à luz das transformações na sociedade internacional. Na medida em que se intensificam os avanços tecnológicos, irão surgir, cada vez mais, casos nos quais os juristas precisarão decidir acerca da eficácia das leis já existentes no contexto da robótica. É preciso, pois, ampliar o debate para desenvolver práticas de responsabilidade adequadas e sem lacunas que possam reduzir a importância da dignidade e dos direitos humanos. Então, faz-se mais do que necessário resolver, de antemão, a aplicabilidade das legislações (tratados e convenções) que regem o Direito Internacional, pois já está evidente que a sociedade globalizada não sabe como lidar com essas novas tecnologias de IA e sistemas de armas autônomas letais que podem potencializar a dimensão das guerras contemporâneas. Contudo, as mudanças ou adaptações da lei internacional dependem de um consenso por parte dos Estados, e esse é, sem dúvidas, o principal desafio para aqueles que não desejam tornar o mundo mais instável e menos pacífico ou seguro. Este é um dos temas mais relevantes da atualidade e suas consequências podem impulsionar novos conflitos ou, ainda, uma nova corrida armamentista por máquinas mais eficazes e com a capacidade de desestabilizar a ordem mundial.

39 Tradução nossa. Texto original: "*Including criminal prosecution, civil damages, disciplinary steps or the offering of redress or compensation. In addition to individual responsibility, institutions, such as States or corporations, may be held accountable*".

CONCLUSÃO

Ao longo da história, as alterações das formas de guerra foram, de fato, muito profundas, principalmente com a deflagração da Primeira e da Segunda Guerras Mundiais. Apesar das terríveis consequências humanas, as guerras foram fundamentais para o desenvolvimento da ciência e da tecnologia. No entanto, atualmente, a criação de armas letais regidas por sistemas autônomos é bastante controversa, uma vez que as guerras justas são tidas como inaceitáveis e a morte de pessoas por máquinas pode minar a dignidade humana, bem como solapar as normas e os princípios do Direito Internacional que regem os conflitos armados. Essa realidade exige, simultaneamente, novas formas de compreensão da sociedade e a adoção de novas alternativas de cooperação entre os Estados, de tal modo que seja possível estabelecer acordos internacionais ou readequar as normas existentes para suprimir as lacunas da lei, seja no tocante à proibição, à restrição ou à autorização, à transparência e à responsabilização jurídica no caso de graves consequências decorrentes do uso dos sistemas de armas autônomas.

As transformações da sociedade internacional das últimas décadas inauguraram um novo momento da história humana. Essa mudança foi impulsionada pelos processos da globalização, pela crescente interdependência e pela formação de fluxos políticos, econômicos e sociais de escala planetária. De fato, o desenvolvimento das novas tecnologias de IA e sua aplicação nas diversas áreas de atuação humana teve um papel primordial. No caso da sociedade internacional globalizada, a convergência desses fatores criou novas possibilidades de cooperação, mas também gerou novas formas de conflitos e novos instrumentos bélicos/militares. Entre esses destacam-se os sistemas de armas autônomas (AWS). A referida ênfase se deve tanto às possíveis e graves violação dos direitos humanos que tais sistemas podem provocar quanto, também, acerca da questão do controle humano significativo sobre o seu uso e a responsabilidade pelos danos ou efeitos colaterais deles decorrentes.

Com efeito, o futuro do desenvolvimento tecnológico de armas bélicas é, sem dúvidas, incerto e, provavelmente, evoluirá de forma contínua. Conforme o exposto ao longo da análise proposta, verifica-se uma verdadeira inadequação dos sistemas de armas autônomas dotados de inteligência artificial para cumprir, integralmente, em um futuro previsível, as conquistas já realizadas pelo Direito Internacional Humanitário. Além disso, pode-se concluir que tanto os seres humanos quanto as tecnologias robóticas de

computadores possuem seus pontos fortes e fracos. Daí, portanto, a necessidade de uma clara regulamentação da utilização dos sistemas de armas autônomas que fortaleçam a parceria e a colaboração entre os seres humanos e as máquinas, de tal modo que seja possível uma certa humanização da guerra e que, mesmo em situações limites, prevaleça a dignidade humana.

Desse modo, é urgente que a sociedade internacional regulamente o tema e estabeleça diretrizes de responsabilidades para todos os envolvidos. Corroborando-se a hipótese inicialmente proposta por este estudo e de acordo com a metodologia de pesquisa aplicada, nota-se que o debate do Direito Internacional acerca desta temática ainda é relativamente pequeno, embora a análise jurídica já tenha situado diversos fatores relevantes para avaliar as possíveis implicações legais. O certo é que as normas internacionais atuais são insuficientes para a regulamentação do tema e, portanto, diante dos novos costumes associados à inteligência artificial e que são empregados na sociedade mundial, é fundamental a produção de um novo tratado internacional regulamentador, no qual os Estados estejam de acordo quanto às diretrizes para utilização dos AWS ou a sua expressa proibição em razão da letalidade desses sistemas. Para isso, serão necessárias a produção de um novo consenso sobre a temática em tela, de acordo com os costumes contemporâneos, e a superação das posições mais contundentes que ou reivindicam a total proibição dos sistemas de armas autônomas ou, então, a sua utilização sem qualquer tipo de regulamentação jurídica internacional. Assim, a iniciativa pressupõe uma capacidade de diálogo e muita lucidez sobre as implicações atuais do uso da inteligência artificial e dos sistemas de armas autônomas em situações de guerra e/ou conflito.

REFERÊNCIAS

BALDASSARRI, Marco Aurélio; NASCIMENTO, Vinícius Damasceno do. Sistemas de armas autônomas e a "guerra justa": a necessidade da vitória *versus* a ética no campo de batalha. *Revista Brasileira de Estudos Estratégicos*, Instituto de Estudos Estratégicos da UFF, v. 11, n. 21, p. 181-213, 2020. Disponível em: http://www.rest.uff.br/index.php/rest/article/viewFile/178/156. Acesso em: 23 ago. 2021.

BECK, Ulrich. *A metamorfose do mundo:* novos conceitos para uma nova realidade. Tradução de Maria Luiza X. de A. Borges. Rio de Janeiro: Zahar, 2018.

_____. *Sociedade de risco*: rumo a uma outra modernidade. Tradução de Sebastião Nascimento. São Paulo: Editora 34, 2011.

BEDIN, Gilmar Antonio. *A sociedade internacional e o século XXI*: em busca da construção de uma ordem mundial justa e solidária. Ijuí: Unijuí, 2001.

_____. *A sociedade internacional clássica*: aspectos históricos e teóricos. Ijuí: Unijuí, 2011.

BHUTA, Nehal; BECK, Susanne; GEIß, R. Present futures: concluding reflections and open questions on autonomous weapons systems. *In*: BHUTA, Nehal C. *et al*. (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

BIRNBACHER, Dieter. Are autonomous weapons systems a threat to human dignity? *In*: BHUTA, Nehal C. *et al*. (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

BOBBIO, Norberto. *O terceiro ausente*: ensaios e discursos sobre a paz e a guerra. São Paulo: Manole, 2009.

CHURCHILL, Winston. *Memórias da Segunda Guerra Mundial.* Volume 1 (1919-1941) e Volume 2 (1941-1945). Tradução de Vera Ribeiro. Rio de Janeiro: HarperCollins, 2017.

CLAUSEWITZ, Carl Von. *Da guerra*. São Paulo: WMF Martins Fontes, 2010.

CORN, Geoffrey S. Autonomous weapons systems: managing the inevitability of "taking the man out of the loop". *In*: BHUTA, Nehal C. *et al*. (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

DUARTE, Érico Esteves. *Tecnologia militar e desenvolvimento econômico:* uma análise histórica. Texto para Discussão, Governo Federal, Secretaria de Assuntos Estratégicos da Presidência da República. Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada (Ipea), 2012. Disponível em: http://repositorio.ipea.gov.br/bitstream/11058/1114/1/TD_1748.pdf. Acesso em: 03 set. 2021.

FEENBERG, Andrew. *Technosystem*: the social life of reason. Cambridge; London: Harvard University Press, 2017.

FERRONI, Marcelo. A evolução das guerras*:* as armas e as táticas em cinco atos, da Antiguidade aos dias de hoje. *São Paulo: O Globo*, Galileu, 2002. Disponível em: http://galileu.globo.com/edic/125/rep_guerra.htm. Acesso em: 02 set. 2021.

GANNON, Megan. Arqueologia revela segredos das trincheiras da Primeira Guerra Mundial. História. São Paulo: *National Geographic Brasil*, 2020. Disponível em: https://www.nationalgeographicbrasil.com/historia/2020/01/arqueologia-revela-segredos-das-trincheiras-da-primeira-guerra-mundial. Acesso em: 01 set. 2021.

GILBERT, Martin. *A Primeira Guerra Mundial*: os 1.590 dias que transformaram o mundo. Tradução de Francisco Paiva Boléo. Rio de Janeiro: Casa da Palavra, 2017.

_____. *A Segunda Guerra Mundial*: os 2.174 dias que mudaram o mundo. Tradução de Ana Luísa Faria e Miguel Serras Pereira. Rio de Janeiro: Casa da Palavra, 2014.

HARARI, Yuval Noah. *Sapiens*: uma breve história da humanidade. Tradução de Janaína Marcoantonio. Porto Alegre: L&MP, 2018.

HARDT, Michael; NEGRI, Antonio. *Multidão*: guerra e democracia na era do Império. Tradução de Clóvis Marques. São Paulo: Record, 2005.

HEYNS, Christof. Autonomous weapons systems: living a dignified life and dying a dignified death. *In*: BHUTA, Nehal C. *et al*. (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

IANNI, Octavio. *A era do globalismo.* Rio de Janeiro: Civilização Brasileira, 2014.

KNUCKEY, Sarah. Autonomous weapons systems and transparency: towards an international dialogue. *In*: BHUTA, Nehal C. *et al*. (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

LAFER, Celso. Declaração Universal dos Direitos Humanos (1948). *In*: MAGNOLI, Demétrio (Org.). *História da paz*: os tratados que desenharam o planeta. São Paulo: Contexto, 2012. p. 247-274.

LEVES, Aline Michele Pedron; BEDIN, Gilmar Antonio. Terrorismo Internacional, Direitos Humanos e Multilateralismo: as (im)possibilidades da paz na sociedade mundial da atualidade. *Inter – Revista de Direito Internacional e Direitos Humanos da UFRJ*, UFRJ, v. 2, n. 1, p. 1-22, 2019. Disponível em: https://revistas.ufrj.br/index.php/inter/article/view/25106. Acesso em: 04 set. 2021.

LÉVY, Pierre. *Cibercultura.* Tradução de Carlos Irineu da Costa. São Paulo: Editor 34, 2010.

LIEBLICH, Eliav; BENVENISTI, Eyal. The obligation to exercise discretion in warfare: why autonomous weapons systems are unlawful. *In*: BHUTA, Nehal C. *et al*. (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

MASSON, Philippe. *A Segunda Guerra Mundial*: história e estratégias. Tradução de Ângela M. S. Corrêa. São Paulo: Contexto, 2017.

MELLO, Celso Albuquerque. *Curso de direito internacional*. São Paulo: Renovar, 1997.

MENEZES, Wagner. *Ordem global e transnormatividade.* Ijuí: Unijuí, 2005.

MOREIRA, Adriano. *Teoria das relações internacionais.* Lisboa: Almedina, 1996.

NYE, Joseph S. The Future of American Power: dominance and decline in perspective. New York: *Foreign Affairs*, november 1, 2010. Disponível em: https://www.foreignaffairs.com/articles/2010-11-01/future-american-power. Acesso em: 03 jan. 2022.

OEA – Organização dos Estados Americanos. *Declaração e Programa de Ação de Viena.* Viena: Conferência Mundial sobre Direitos Humanos – OEA, 1993.

PAGALLO, Ugo. *The Laws of Robots*: Crimes, Contracts, and Torts. Cham: Springer, 2013.

PÉREZ LUÑO, Antonio Enrique. Sistemas expertos jurídicos: premisas para un balance. *In*: PEÑA, Lorenzo *et al.* (Orgs.). *Calculemos...* Matemáticas y libertad. Espanha: Trotta, 1996.

PHILLIPS, G. Longbow and hackbutt: weapons technology and technology transfer in early modern England. *Technology and Culture*, Johns Hopkins University Press, v. 40, n. 3, p. 576-593, 1999. Disponível em: www.jstor.org/stable/25147360. Acesso em: 03 set. 2021.

ROFF, Heather M.; MOYES, Richard. Meaningful Human Control, Artificial Intelligence and Autonomous Weapons. *In*: *Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons*, Geneva, Switzerland, 2016. Disponível em: https://article36.org/wp-content/uploads/2016/04/MHC-AI-and-AWS-FINAL.pdf. Acesso em: 05 set. 2021.

SCHNEEWIND, Jerome B. *A invenção da autonomia*. São Leopoldo: Unisinos, 2001.

SHARKEY, Noel. Staying in the loop: human supervisory control of Weapons. *In*: BHUTA, Nehal C. *et al.* (Orgs.). *Autonomous Weapon Systems*: Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

SLAUGHTER, Anne-Marie. *The Chessboard and the Web*: Strategies of Connection in a Networked World. New Haven: Yale University Press, 2017.

STOP KILLER ROBOTS. *The Problem.* Nova York: Human Rights Watch, 2021. Disponível em: https://www.stopkillerrobots.org/learn/#problem. Acesso em: 05 set. 2021.

TAMBURRINI, Guglielmo. On banning autonomous weapons systems: from deontological to wide consequentialist reasons. *In*: BHUTA, Nehal C. *et al.* (Orgs.). *Autonomous Weapon Systems:* Law, Ethics, Policy. Cambridge: Cambridge University Press, 2016.

TOTA, Pedro. Segunda Guerra Mundial. *In*: MAGNOLI, Demétrio (Org.). *História das guerras.* São Paulo: Contexto, 2011. p. 355-389.

TRINDADE, Rodrigo. Guerra 2.0, o futuro chegou. *São Paulo: UOL*, Tilt, 2019. Disponível em: https://www.uol.com.br/tilt/reportagens-especiais/novas-tecnologias-irao-moldar-guerra-do-amanha. Acesso em: 04 set. 2021.

WISCHMEYER, Thomas; RADEMACHER, Timo (Eds.). *Regulating Artificial Intelligence.* Cham: Springer, 2020.

ZOLO, Danilo. *Terrorismo Humanitario*: de la guerra del Golfo a la carnicería de Gaza. Traducción de Juan Vivanco Gefaell. Barcelona: Edicions Bellaterra, 2009.

Sobre o autor e sobre as autoras:

Gilmar Antonio Bedin | *E-mail:* gilmarb@unijui.edu.br

Doutor e Mestre em Direito pela Universidade Federal de Santa Catarina (UFSC). Pós-Doutorando pela Universidade de Santiago de Chile (USACH). Bacharel em Direito pela Universidade de Santa Cruz do Sul (Unisc). Professor dos Cursos de Graduação em Direito e dos Programas de Pós-Graduação *Stricto Sensu* em Direito – Mestrado e Doutorado – da Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí) e da Universidade Regional Integrada do Alto Uruguai e das Missões (URI). Líder do Grupo de Pesquisa do CNPq: Direitos Humanos, Governança e Democracia (*Mundus*).

Aline Michele Pedron Leves | *E-mail:* aline.leves@sou.unijui.edu.br

Doutoranda e Mestra pelo Programa de Pós-Graduação *Stricto Sensu* em Direito – Mestrado e Doutorado em Direitos Humanos – da Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí). Bacharela em Direito pela Unijuí. Bolsista de Doutorado da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Pesquisadora integrante do Grupo de Pesquisa do CNPq: Direitos Humanos, Governança e Democracia (*Mundus*). Advogada (OAB/RS).

Laura Mallmann Marcht | *E-mail:* laura.marcht@sou.unijui.edu.br

Mestranda pelo Programa de Pós-Graduação *Stricto Sensu* em Direito – Mestrado em Direitos Humanos – da Universidade Regional do Noroeste do Estado do Rio Grande do Sul (Unijuí). Bacharela em Direito pela Unijuí. Bolsista de Mestrado da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes). Pesquisadora integrante do Grupo de Pesquisa do CNPq: Direitos Humanos, Governança e Democracia (*Mundus*).

Data de Submissão: 30 de setembro de 2021.
Data de Aceite: 10 de janeiro de 2022.

A Proteção Jurídica Fundamental da Confidencialidade e da Integridade dos Sistemas Técnicos de Informação de Uso Próprio[1]

The Fundamental Legal Protection of Confidentiality and Integrity of Informational Technical Systems of Own Use

PROF. DR. WOLFGANG HOFFMANN-RIEM[2]
Bucerius Law School.

Italo Roberto Fuhrmann (Trad.)[3]

Jacqueline de Souza Abreu (Rev. Técnica)[4]

RESUMO: O artigo insere a decisão do Tribunal Constitucional Federal da Alemanha sobre a assim chamada busca online (Online-Durchsuchung) no contexto mais abrangente do desenvolvimento das tecnologias da informação e da sua utilização desde o julgamento do caso da Lei do Recenseamento, bem como elabora as especificidades do novo direito fundamental.

ABSTRACT: The article inserts the German constitutional court decision on the so-called Online Search in the broader context of the development of information technology and its use since the Census Case, as well as elaborates on the specifics of the new fundamental right.

SUMÁRIO: I – O início da jurisprudência do censo e o seu significado contínuo; II – Mudanças dos riscos e das oportunidades por meio das tecnologias da comunicação; II.1 Oportunidades e riscos; II.2 Necessidades e possibilidades de proteção; III – Em especial: proteção da confiança nos

1 O artigo é a versão revisada e ampliada de uma palestra na conferência de proteção de dados da Friedrich-Ebert-Stiftung, realizada em Berlim, em 1º de julho de 2008. Os agradecimentos por valiosas sugestões vão para Marion Albers, Matthias Bäcker e Ulf Buermeyer.

2 Orcid: https://orcid.org/0000-0003-1085-6673.

3 Orcid: https://orcid.org/0000-0002-3914-8200.

4 Orcid: https://orcid.org/0000-0003-0450-4102.

sistemas técnicos de informação de uso próprio; III.1 Proteção da confiança; III.2 Relevância da personalidade; IV – Abordagens para a proteção de direitos fundamentais; IV.1 Defesa e proteção; IV.2 Normas relevantes de direitos fundamentais; IV.2.a Sigilo das telecomunicações; IV.2.b Proteção do domicílio; IV.2.c Direito fundamental à proteção da personalidade; V – Em especial: a garantia jurídico-fundamental da confidencialidade e da integridade dos sistemas técnicos de informação de uso próprio; V.1 O ponto de partida; V.2 Novas dimensões da necessidade protetiva; V.2.a Amplitude da intervenção relativa à personalidade; V.2.b Sobre dados gerados pelo sistema; V.2.c Criação de novas imagens da personalidade de nova profundidade e amplitude; V.2.d Risco de falsificação de dados; V.2.e Neutralização das possibilidades de autoproteção; V.2.f Possibilidade de acesso de terceiros; V.2.g Grande variação de pessoas envolvidas; V.3 Esclarecimento da peculiaridade da situação de risco e da respectiva proteção de direitos fundamentais pelo Tribunal Constitucional Federal da Alemanha; V.3.a Diferenciação ao nível do âmbito de proteção; V.3.b Reação à especial qualidade do risco; V.3.c Delimitação com relação ao direito à autodeterminação informativa; V.3.d Necessidade de outras concretizações; VI – Limites aos direitos fundamentais; VI.1 Requisitos jurídicos materiais e processuais; VI.2 Núcleo da vida privada; VII – Concorrência com outras normas de direitos fundamentais; VII.1 Intervenção no âmbito do domicílio; VII.2 Concorrência com o sigilo de telecomunicação especialmente com as TKU-Fonte; Conclusão; Posfácio.

I – O INÍCIO DA JURISPRUDÊNCIA DO CENSO E O SEU SIGNIFICADO CONTÍNUO

15 de dezembro de 1983 – há um quarto de século – foi um grande dia para a ampliação da proteção dos direitos fundamentais na Alemanha: nesta data, foi proferida a decisão do censo pelo Tribunal Constitucional Federal da Alemanha[5]. Neste contexto, foi reconhecido o "direito à autodeterminação informativa" como inerente à proteção da personalidade pelo Tribunal Constitucional. Qualquer pessoa que leia hoje a decisão deve ficar surpreendida com o fato de que um recenseamento da população, isto é, de uma coleta estatística de informações como nome, endereço, meio de sustento, profissão e outros dados[6] análogos, poderia causar uma celeuma tão grande e, ao mesmo tempo, estimular uma decisão tão inovadora.

5 BVerfGE 65, 1.

6 No julgamento do censo, o Tribunal Constitucional usou o termo "dados" em um contexto para o qual a literatura de tecnologia da informação usa parcialmente o termo "informação"; sobre isso, ver, por todos, Albers, Informationelle Selbstbestimmung, 2005, p. 87 e ss.; Vesting, in: Hoffmann-Riem/Schmidt-Aßmann/Vosskuhle (Org.), Grundlagen des Verwaltungsrechts vol. 2, II (2008), § 20 número de margem 11 e ss. A seguir, o termo dados continuará a ser utilizado de acordo com o exercício na literatura jurídica, mesmo na medida em que o termo informação seja utilizado na literatura de tecnologia da informação/ciência. Dados são caracteres objetivados, informações com conteúdos significativos formados pelos destinatários ou em sistemas de comunicação (Cf., por exemplo, Albers, op. cit., p. 141 e ss.). A proteção de dados é uma dimensão protetora do direito fundamental à autodeterminação informativa, bem como o direito fundamental à proteção da confidencialidade e da integridade dos próprios sistemas de tecnologia da informação, porém, em última análise, os dados são protegidos em virtude da informação transportado com ele.

Nesta senda, pode ter desempenhado um papel a circunstância de que o ano de 1984, o mesmo que Orwell tinha escolhido como título de seu livro numa perspectiva futurista acerca dos perigos do Estado de vigilância através do "Big Brother", estava por vir e inspirou as fantasias. Naquele tempo, a tecnologia computacional ainda estava no seu início. O processamento de dados era realizado em grande medida por grandes computadores centrais, que eram pesados e volumosos, caros e, sobretudo, lentos e com pouca capacidade de armazenamento se comparados aos dias de hoje. Um computador com boa capacidade funcional e de preço acessível para todos – como hoje o PC – estava apenas no começo do desenvolvimento, e as utilizações, tal como hoje o smartphone possibilita, eram, quando muito, objeto de fantasia. Para o armazenamento de dados, que hoje um pendrive pode realizar, eram necessárias grandes máquinas imóveis. Por conseguinte, a consciência ainda não tinha se formado acerca das grandes oportunidades comunicativas de ação que o computador viria a possibilitar nos próximos anos, especialmente graças à rede internacional também de computadores privados e ao desenvolvimento da internet. O que estava em primeiro plano na discussão pública não eram tais oportunidades, mas as ameaças à liberdade por meio do levantamento de dados e seu processamento pelo Estado. Embora os dados levantados através do censo devessem permanecer anônimos, foi discutido o risco de sua individualização e o perigo associado de sua utilização indevida.

Na dogmática dos direitos fundamentais, tratava-se de uma ativação da proteção de um direito fundamental de defesa, que primeiro necessitaria ganhar contornos. O Tribunal Constitucional Federal da Alemanha conseguiu resumir, em poucas palavras, a reflexão fundamental, que, neste meio-tempo, não perdeu em nada do seu significado[7]: o direito geral da personalidade garantido no art. 2º, § 1º, c/c art. 1º, § 1º, da LF[8] poderia

> igualmente ganhar significado em relação aos modernos desenvolvimentos e às novas ameaças à personalidade humana a eles associadas. As concretizações realizadas até então pela jurisprudência não descreviam de forma conclusiva o conteúdo do direito de personalidade. Ele abrange também a

7 As seguintes citações são derivadas do BVerfGE 65, 1, 41-44 (supressões não são registradas).

8 Se é necessário e adequado usar também o art. 1º, § 1º, da Lei Fundamental como uma justificativa para o direito geral de personalidade, faz-se referência aos argumentos consideráveis em Britz, Freie Entfaltung durch Self-Presentation, 2007, em especial p. 25 e ss. Em qualquer caso, deve ser necessário derivar a proteção de dados do art. 2º, § 1º, da Lei Fundamental, na medida em que a proteção da personalidade relacionada com a dignidade humana não seja afetada, a menos que outros direitos fundamentais – como os arts. 12 e 14 da Lei Fundamental – não sejam aplicáveis.

> prerrogativa do indivíduo, decorrente das reflexões da autodeterminação, de decidir por si mesmo quando e dentro de quais limites os fatos pessoais da vida são revelados. Este direito à "autodeterminação informativa" não é garantido sem limites. O indivíduo não tem um direito no sentido de um domínio absoluto e ilimitado sobre "seus" dados; ele é, ao contrário, uma personalidade que se desenvolve dentro da comunidade social e que depende da comunicação.

A dogmática da decisão baseada nos direitos de defesa classificou a atividade do Estado como uma intervenção numa posição jurídica individual do sujeito, que, em última análise, funcionou como um direito a um dado próprio[9], cuja divulgação e utilização o indivíduo deveria poder livremente decidir – ainda que fosse reconhecida uma vinculação a contextos sociais. Uma vez que se tratava da defesa contra uma intervenção estatal, a decisão se assentou na relação entre o Estado e o cidadão.

Tendo em vista a necessidade aparentemente ilimitada de informações do Estado, ainda há aqui uma necessidade contínua de proteção. Isto é demonstrado, por exemplo, pelas muitas autorizações de acesso aos dados no interesse de garantir a segurança pública e a persecução criminal, que são encontrados em número crescente nas leis policiais e de proteção da constituição, bem como nas normas de processo penal e, em particular, as que foram criadas como resultado do 11.09.2001 como um meio (também) de combate ao terrorismo. O fato de que uma série dessas autorizações, ou pelo menos o tratamento delas no caso concreto, foram vistas nos últimos anos como inconstitucionais pelo BVerfG – e, em especial, como uma violação do direito fundamental à autodeterminação informativa[10] – sinaliza a contínua importância da proteção dos direitos de defesa.

---

9 De modo crítico, por exemplo, Hoffmann-Riem AöR 123 (1998), 513, 520 f.m.w. Indicações. Crítica fundamental da construção do Tribunal Constitucional, especialmente em Albers (nota 2), por exemplo, p. 238 e passim. A ideia do direito aos próprios dados é ainda mais inadequada quando se trata de dados sobre informações relativas ao comportamento de várias pessoas sem que os interesses de uma das pessoas em causa sejam juridicamente dignos de proteção. Torna-se ainda mais difícil com a atribuição de um dado a uma pessoa se o seu valor informativo deriva da combinação com outros dados que são ou foram gerados por outras pessoas.

10 Cf., em especial, BVerfGE 115, 320 e ss. (busca computadorizada) e BVerfG, acórdão de 11 de maio de 2008 – 1 BvR 2074/05, 1 BvR 1254/07 = NJW 2008, 1505 e ss. (registro automático de placas).

II – MUDANÇAS DOS RISCOS E DAS OPORTUNIDADES POR MEIO DAS TECNOLOGIAS DA COMUNICAÇÃO

II.1 Oportunidades e riscos

Em comparação com a época do julgamento do caso do censo, o nível de risco da situação mudou intensamente, assim como aumentaram enormemente as oportunidades de uso comunicativo da eletrônica para o desenvolvimento individual e coletivo. Praticamente qualquer pessoa hoje tem acesso a computadores com boa capacidade; cerca de 35 milhões de alemães utilizam o sistema global de rede da internet. Em julho de 2008, havia no mundo mais de 860 milhões de usuários de internet.

O que é característico não é mais o armazenamento centralizado de dados, mas sim usos descentralizados e a rede de sistemas de computação altamente eficientes muitas vezes de acesso global. A digitalização em conjunto com a tecnologia computacional – também em vista da globalização – trouxe uma evolução comunicativa, cuja importância para o desenvolvimento social não é inferior à da revolução industrial do século XIX. Computadores grandes e pequenos e as respectivas infraestruturas comunicativas das técnicas de informação se tornaram em forças produtivas centrais em praticamente todas as esferas da vida, seja para o desenvolvimento da vida privada, seja para o cumprimento de tarefas por parte do Estado e da economia por parte de empresas. As tecnologias de comunicação moldam o exercício real dos direitos fundamentais em praticamente todos os ambientes sociais[11]. Muitas das tecnologias e dos serviços eram desconhecidos à época da decisão do censo, por exemplo, ISDN, RFID, WLAN, UMTS; serviços como o e-commerce, o governo eletrônico, os sistemas de navegação; as redes sociais, como StudiVZ, comunidades virtuais como o Second Life, ou mesmo os métodos de investigação como o kfz-Scanning ou a busca on-line.

O Estado é apenas, de forma limitada, o promotor e garante da capacidade funcional das infraestruturas técnicas de informação, de resto, e mesmo principalmente, são as empresas privadas, incluindo aquelas com poder global – como Google, Microsoft ou as grandes empresas de telecomuni-

11 Sobre a computação ubíqua (o processamento de dados ubíquo), ver, por todos, Kühling Die Verwaltung 40 (2007), 153 e ss. Da mesma forma, as contribuições em Roßnagel/Sommerlatte/Wienand (Org.), Digitale Visionen – zur Gestaltung ubiquitous information technologies, 2008 e em Mattern (Org.), The Informatization of Everyday Life – Living in Smart Environments, 2007.

cações. Entre as várias empresas, e em relação ao cidadão, mas também na relação entre as empresas e o Estado, existem consideráveis assimetrias de poder. O fato de que o uso do poder, e, portanto, os riscos do abuso de poder não estão de forma alguma limitados ao Estado, está se tornando cada vez mais aparente para o público, por exemplo, quando é discutida a profusão de dados e possibilidades de seleção que o Google, por exemplo, tem à sua disposição[12], ou quando escândalos são descobertos, como o uso dos dados de conexão dos clientes da Deutsche Telekom para a vigilância dos próprios funcionários[13], ou a venda ilegal de dados bancários[14]. Ao mesmo tempo, no entanto, o Estado tem acesso aos dados, especialmente no âmbito da prevenção e combate a perigos, bem como da persecução penal, de modo que a proteção contra tais violações também deva ser garantida.

II.2 Necessidades e possibilidades de proteção

Os direitos fundamentais, em particular a proteção dos direitos da personalidade, a liberdade de comunicação e a proteção da residência, são orientados para a proteção contra intervenções do Estado, mas também contra intervenções da liberdade através de privados. Em especial, a liberdade de desenvolvimento comunicativo está afetada. A liberdade de comunicação é uma liberdade que é usada em combinação com outras liberdades[15]. A este respeito, a posição individual só pode ser descrita a partir da relação social. Um pensamento dogmático dos direitos fundamentais centrado no indivíduo "solitário" não poderia captar adequadamente as dimensões sociais da liberdade de comunicação e, portanto, as exigências de proteção a ela relacionadas.

Na medida em que a comunicação pessoal é estruturada em termos tecnológicos, esta requer uma proteção de direitos fundamentais efetiva também no que se refere à proteção da infraestrutura tecnológica da comunicação e da sua utilização concreta, uma vez que isso está relacionado com a liberdade do indivíduo. A capacidade funcional não tem apenas um lado técnico, mas também um lado social, que pode ser influenciado de forma normativa, por exemplo, assegurando a liberdade de acesso, a liber-

12 Cf., por todos, Maurer Informatik Spektrum 30 (2007), 273 e ss.

13 Acerca do assim chamado escândalo da Telekom, v. Süddeutsche Zeitung em 29.05.2008, p. 2, assim como do dia 30.05.2008, p. 1, e Scherer MMR 2008, 433 e ss.

14 Cf., por todos, Dams Die Welt em 18.08.2008.

15 Para obter informações gerais sobre este conceito, ver, em especial, Suhr, Entfaltung des Menschen durch die Menschen, 1976; (Org.), EuGRZ 1984, 529, 537. Cf. igualmente Albers (nota 2) e Britz (nota 4), p. 45 e ss.

dade de manipulação, e geralmente através de proteção contra o uso ou mesmo abuso de poder. As diversas dimensões da capacidade funcional referem-se a diferentes potenciais de perigo e vulnerabilidades, assim como a diferentes atores que salvaguardam ou põem em perigo a capacidade funcional. Por isso são necessários conceitos multipolares e multidimensionais de proteção da liberdade.

O Estado só pode assegurar de maneira limitada o funcionamento das infraestruturas de comunicação – não apenas por conta do alcance global da rede, mas também pelo domínio de atores privados na criação e manutenção das redes mediante prestação de serviços. Também se situam, neste contexto, os atores que estabelecem suas próprias leis (como, por exemplo, ICANN)[16]. Nada obstante, o Estado pode utilizar o seu poder regulamentar dentro do âmbito das normas por ele estabelecido e ampliado, eventualmente, por atos jurídicos interestatais.

A asseguração das condições reais da liberdade de conduta, em especial da liberdade comunicativa em relações conectadas tecnologicamente, não pode ser alcançada tão somente por meio do controle normativo estatal (ou privado) da conduta, mas igualmente por meio de outras medidas[17], como, por exemplo, através de exigências legais que afetem a forma como o sistema de comunicação é configurado, ou que possibilitem a proteção de dados e a autoproteção tecnológica, por exemplo, através da criptografia. Neste contexto, o Estado pode conceder incentivos, se necessário também por meio de proibições e imperativos que podem levar à ativação de funções distintas de proteção. Isto porque a proteção de dados direcionada, tanto quanto possível, para o paradigma da autodeterminação, e não apenas na definição do objetivo de proteção, mas também nas precauções de proteção, atinge seus limites factuais – e normativos – onde o indivíduo carece dos meios de proteção ou conscientização da necessidade de proteção de seus dados pessoais[18].

16 A "Internet Corporation for Assigned Names and Numbers", com sede em Marina del Rey (Califórnia), administra as principais estruturas da internet, nomeadamente, entre outras, a atribuição de blocos de endereços IP (os chamados espaços de endereço) e o servidor DNS central, que atua como uma "lista telefônica" da internet, convertendo as informações de endereço textual (por exemplo, www.bundesverfassungsgericht.de) em endereços IP (no exemplo: 134,96. 83. 81).

17 Cf. Albers (nota de rodapé 2), por exemplo, p. 466 e ss., 544 e ss.

18 A diferença entre a motivação das massas que se opunham ao censo dos anos oitenta, por um lado, e a vontade que hoje é generalizada de divulgar até os detalhes mais privados nas redes sociais (como alunos/StudiVZ), nos programas de fidelização (por exemplo, Payback) e portais de internet, além disso, limita as chances de realização de auxílio estatal para autoproteção, mas não o torna dispensável como uma oferta – ver também abaixo V 2 d.

III – EM ESPECIAL: PROTEÇÃO DA CONFIANÇA NOS SISTEMAS TÉCNICOS DE INFORMAÇÃO DE USO PRÓPRIO

Atualmente, pode-se observar que as infraestruturas técnicas da informação[19] e comunicação coletam e processam[20] cada vez mais dados pessoais, mesmo em computadores de uso próprio, que são utilizados, por exemplo, como arquivo para informações a serem retidas, como um auxílio ao realizar suas próprias tarefas (escrita, aritmética, de gestão), como meio de entretenimento (jogos de computador, biblioteca digital, biblioteca de áudio, biblioteca de vídeo) ou para controle (remoto) de sistemas de gestão doméstica em "casas inteligentes"[21], bem como para criar os chamados veículos inteligentes[22] ("Internet das Coisas")[23]. No uso online, o computador está conectado em rede com outros computadores, e os dados contidos ou gerados nele podem ser eventualmente utilizados em outros computadores. A integração em redes, em particular na internet global, permite, para além do acesso aos dados lá disponíveis, os serviços oferecidos ao trabalhar com o próprio computador, embora muitas vezes não seja conhecido do usuário qual software ainda está "a seu serviço" ou o que é utilizado para acesso às suas informações. Se – como é de se esperar – cada vez mais no futuro crescerem juntos os usos fornecidos com o software no computador utilizado (a chamada "cloud computing"[24] ou "services in the cloud"), sua propagação e diversidade, assim como a falta de controle, continuarão aumentando para

19 Para as perspectivas de desenvolvimento, ver, por todos, Roßnagel, Datenschutz im IT, 2007, especialmente p. 26 e ss. Sobre a necessidade de uma estrutura transdisciplinar, ver Rolf, Mikropolis 2010, 2008, p. 95 e ss.

20 Cf., por todos, Kutscha NJW 2008, 1042, 1044.

21 Em sua decisão sobre buscas online, o julgamento de 27.02.2008 – 1 BvR 370/07, 1 BvR 595/07 = NJW 2008, 822 (disponível em: www.bundesverfassungsgericht.de), número de margem 202, já mencionado – para uma visão prática: Bayerlein-Hoppe Elektrobörse Handel 02/2004, 12 e ss. Certamente não é por acaso que a exposição internacional de rádio em Berlim fez uma reorientação conceitual em 2008, de modo que as precauções eletronicamente estruturadas para "casas inteligentes" agora também foram integradas em uma feira de comunicações.

22 Ver, especialmente, a iniciativa da União Europeia apresentada, por exemplo, no comunicado da Comissão "Para uma mobilidade mais segura, mais limpa e mais eficiente no âmbito europeu: primeiro relatório sobre a iniciativa 'Veículo Inteligente'", COM (2007), 541. Os sistemas de transporte inteligentes destinam-se, em particular, a aumentar a segurança do tráfego e a eficiência energética e a proporcionar uma maior utilização de tecnologias de informação e comunicação, que, ao mesmo tempo, transferem informações de veículo para veículo, entre veículo e infraestrutura, de veículo para sistemas de chamada de emergência (incluindo precauções com controle de localização exata). Veja também Dencker zfs 2008, 423 ff.; Vieweg, in: 45. VGT, 2007, p. 292 e ss.

23 Ullinger/ten Hompel (Org.), Internet der Dinge, 2007; Fleisch/Mattern (Org.), Das Internet der Dinge – Ubiquitous Computing und RFID in der Praxis, 2005. Cf. também acima na nota de rodapé 7.

24 Sobre o tema, cf. David Chappell, A Short Introduction to Cloud Platforms. An enterprise-oriented view, 2008, disponível em: www.davidchappell.com. A "nuvem" representa uma metáfora para as infraestruturas complexas opacas e em constante movimento que a comunicação baseada em rede acessa sem que o usuário saiba ou mesmo seja capaz de controlá-la.

o usuário. A perda de controle é inevitável. Princípios normativos como minimização de dados e prevenção de dados (Seção 3a (1) BDSG) não se tornarão supérfluos, mas perderão parte de sua eficácia, se a infraestrutura de rede for acessada – o que é praticamente inevitável com o uso online.

O disco rígido de muitos PCs já oferece uma imagem espelhada dos interesses e das inclinações pessoais, da situação econômica, assim como do bem-estar físico e psicológico, ou mesmo do comportamento de seus usuários[25]. Porém, as informações confidenciais não estão apenas no seu próprio computador "armazenadas", mas também estão localizadas na própria rede. Quem tem acesso ao sistema de tecnologia da informação pode, em certa medida – pelo menos em parte – ter acesso ao "cérebro externalizado"[26] ou mesmo à "psique externalizada", mas também a muitas outras informações importantes da personalidade afetada. Essa "vulnerabilidade" do direito da personalidade leva a demandas por proteção, inclusive (ainda) no que diz respeito aos dados e coletas de dados a princípio conhecidos pelo usuário, mas também em relação aos dados de conteúdo gerados durante o processo de uso, bem como aos dados voláteis ou gerados permanentemente (os dados funcionais) e suas possibilidades de uso, muitas vezes não conhecidos pelo usuário. As garantias ganham relevância em termos de direitos fundamentais, na medida em que sejam necessárias para proteger a personalidade (relevância da personalidade).

III.1 Proteção da confiança

Uma proteção eficaz de tais dados e da comunicação divulgada por eles não requer apenas a proteção contra acesso, mas também abrange a confiança[27] de que o hardware e o software utilizados e, no geral, as infraestruturas comunicacionais técnicas da informação funcionem não só em termos técnicos, mas igualmente nos contextos de aplicação[28], assim como

25 Assim, por exemplo, Kutscha NJW 2008, 1043.

26 Hassemer Süddeutsche Zeitung, aos 11.06.2008, formulou: "O computador é uma parte externalizada do corpo".

27 Basicamente sobre confiança e suas dimensões, ver as contribuições de Klumpp et al. (Ed.), Informationelles trust for the information society, 2008. As muitas facetas do conceito de confiança e as teorias sobre a construção da confiança não podem ser tratadas aqui. Para diferentes perspectivas disciplinares, consulte Möllering, in: Max-Planck Institute for Social Research, anuário 2007/2008, p. 73 e ss.

28 A este respeito, trata-se também de proteção funcional, ver Hornung CR 2008, 299, 302. Do ponto de vista dos direitos fundamentais, a proteção funcional deve ser entendida como um meio de proteger a privacidade pessoal.

o usuário pode esperar[29], de modo que ele possa, nesse contexto, confiar na proteção de dados técnicos de informação armazenados ou comunicados (confiança relacionada ao sistema). Às expectativas protegidas normativamente relacionadas à confiança pertence a confidencialidade fundamental do próprio sistema de tecnologia da informação[30], que é a própria base da confidencialidade da comunicação, portanto uma proteção em face do acesso do Estado ou de terceiros[31]. As expectativas de proteção também incluem a integridade do sistema de tecnologia da informação, ou seja, a proteção contra a superação de obstáculos que protegem contra intrusões, bem como contra avarias e manipulações[32], por exemplo, contra falsificações, contra complementações por meio de outros dados ou por meio de softwares que podem manipular o manuseio de dados[33]. Também há necessidade de proteção contra manipulação do hardware utilizado, bem como contra a infiltração e manipulação dos programas que (como o sistema operacional ou o software do usuário) habilitam a funcionalidade ou oferecem acesso a terceiros ao sistema.

O Tribunal Constitucional Federal da Alemanha utiliza o conceito dos sistemas técnicos de informação como juridicamente constitucional, cujos contornos ainda precisam ser desenvolvidos, e que, em hipótese alguma, pode ser extraído de forma isolada da literatura técnica informacional. Isto deixa claro que as necessidades de proteção especial só se aplicam a sistemas complexos de tecnologia da informação, o que não ocorre com os dispositivos de controle eletrônicos que não estão em rede de tecnologia doméstica[34], mas sim com computadores pessoais em rede, telefones celulares mais complexos ou assistentes digitais pessoais (PDAs)[35]. Os requisitos de complexidade suficiente também podem ser preenchidos por um pen-

29 Cf. Volkmann DVBl. 2008, 590, 592.

30 O termo sistemas de tecnologia da informação ainda não foi definido em termos legais. O Tribunal Constitucional (nota 17) retomou-o da literatura de informática, cuja terminologia constava do regulamento jurídico impugnado na decisão em questão. Uma das futuras tarefas dogmáticas será descrever as estruturas juridicamente relevantes dos "próprios sistemas de tecnologia da informação" em mais detalhes, de uma forma que seja voltada para a relevância pessoal da proteção do sistema. Mesmo quando o Tribunal Constitucional fala em sistemas de tecnologia da informação "próprios" (ou melhor: autoutilizados), o contexto deixa claro que a proteção da personalidade é um ponto de referência decisivo para a proteção do sistema de direitos fundamentais.

31 Cf., igualmente, Britz (nota de rodapé 4), p. 77.

32 Trata-se, neste sentido, na terminologia técnica da informática também de "Security" no sentido da segurança de TI. Sobre o tema, cf. Kubicek, in: Klumpp u. a. (nota de rodapé 23), p. 17, 25 e ss.

33 Acerca da proteção da "exatidão informativa", cf. Albers (nota de rodapé 2), p. 119 e ss.; Britz (nota de rodapé 4), p. 52 e ss.

34 Cf. BVerfGE (nota de rodapé 17), número de margem 202.

35 Cf. BVerfGE (nota de rodapé 17), número de margem 194.

drive conectado ao computador ou por um disco rígido externo conectado a ele[36].

III.2 Relevância da personalidade

No entanto, o sistema técnico de informação não é protegido em termos de direitos fundamentais por sua própria vontade[37], mas apenas na medida em que sua confidencialidade e integridade tenham relevância para a personalidade[38]. Isto, por sua vez, resulta do tipo de dados que são transportados com a ajuda do sistema ou que são ou podem ser armazenados nele. A proteção de dados pretendida por meio da proteção dos sistemas técnicos de informação também se estende aos dados de relevância para a personalidade que são mantidos na memória de trabalho e armazenados temporária ou permanentemente na mídia de armazenamento do sistema (possivelmente apenas de forma indireta)[39].

Entretanto, como o usuário em regra não sabe, e nem tem como saber, quais dados pessoais ou eventualmente quais dados relativos a sua personalidade nos atuais e complexos sistemas técnicos de informação são gerados durante o processo de uso, onde e por quanto tempo são registrados (armazenados), e em quais contextos de uso e por quem são utilizados, ele praticamente não pode exercer o seu direito de autodeterminação sobre a divulgação e a utilização de tais dados – com o qual ele até agora decidiu em que medida ele poderia confiar no sigilo. A possibilidade de disposição autônoma também deixa de existir quando houver ciência da natureza dos dados, se ele estiver sobrecarregado pela autoproteção ou a autoproteção levar a perdas funcionais inaceitáveis. O ganho em possibilidades técnicas

36 Sobre o tema, em mais detalhes, Bäcker, in: Brink/Rensen (Org.), Aktuelle Rechtsprechung des Bundesverfassungsgerichts, 2009 (na edição), sob III, 2a; Böckenförde JZ 2008, 925, 929 nota de rodapé 41.

37 Neste aspecto, existem proteções complementares através de outros direitos fundamentais, como art. 12, 14, da LF. Um conceito de sistema técnico de informação baseado na proteção da propriedade ainda necessita elaboração.

38 O risco temido por Eifert NVwZ 2008, 521, 522, de que a proteção da integridade transforme a proteção dos direitos fundamentais em direito fundamental não pessoal e tecnológico, não existe se, como alega o Tribunal Constitucional, a relação de proteção de integridade ao direito fundamental à garantia da personalidade do art. 2º, § 1º c/c o art. 1º, § 1º, da LF seja preservada, mesmo que se estenda ao nível de perigo pessoal. O componente relacionado à proteção pessoal também prevê Lepsius, in: Roggan (Ed.), Online-Durchsuchungen, 2008, p. 21, 32 e ss. Quando ele descreve a nova dimensão da proteção como a "proteção desindividualizada da funcionalidade" desses sistemas e a referência de personalidade exigida pelo tribunal apenas como "área de proteção de contorno, mas não individualizante" (Op. cit., p. 35).

39 Uma visão ilustrativa do potencial de perigo e as possibilidades de acesso estatal encoberto aos sistemas de computador são dadas por Buermeyer HRRS 2007, 154 e ss.

de intercâmbio de informações corresponde a uma perda estrutural de autonomia informacional[40]. No entanto, a proteção funcional relacionada ao sistema permite – dentro de certos limites – precauções para a compensação desta perda de autonomia, porém dificilmente para a restauração da possibilidade de decisão autodeterminada sobre o tratamento dos seus próprios dados.

O paradigma[41] de assegurar a liberdade por meio da possibilidade fundamental de tomar decisões autônomas sobre acesso e uso[42] de dados, no qual o Tribunal Constitucional Federal baseia o direito de autodeterminação informacional, indica inicialmente um objetivo de proteção da liberdade, mas também se refere a possíveis formas de alcançar o objetivo por meio da autodeterminação. O direito de proteção de dados assumiu isso através de certos instrumentos, como a função do consentimento (§ 4 para. 1, § 4a BDSG), ou o pedido para utilizar as possibilidades de anonimização e pseudonimização (§ 3a BDSG). Na referência a tais instrumentos, entretanto, a proteção da personalidade se baseia em premissas empíricas que estão cada vez mais em erosão devido ao desenvolvimento da tecnologia informática, de constelações de redes e de muitos novos serviços. O direito de proteção de dados deve lidar com isso. Isto tem – apenas para citar um exemplo – consequências para a relevância da exigência do consentimento[43]. Para aquele que não pode ver sobre o que está consentindo – que não pode saber quem, o quê[44], quando e em que ocasião sobre quem sabe –, não pode autorizar outros a processar dados de maneira "informada"[45] e, portanto, autodeterminada; sem um fundamento suficiente de informação, o consentimento é reduzido a uma fórmula sem força de legitimação material ou se torna até mesmo uma ficção. Uma proteção de dados eficaz só pode se basear na possibilidade de proteção da liberdade pelos próprios titulares dos dados na medida em que estes possam exercer efetivamente esta possibilidade. Além disso, há a necessidade de mecanismos de proteção suplementares. Muitos esforços foram feitos no passado para estabelecer

40 De modo resumido, in: Sokol (Org.), Persönlichkeit im Netz: Sicherheit – Kontrolle – Transparenz, 2007, p. 4 e ss.

41 Sobre ele, v. BVerfGE 65, 1, 42 e ss.

42 Que essa competência não pode ser (mal)entendida como quase direito à propriedade, já foi acentuado (nota de rodapé 7).

43 Sobre ele, cf., por todos, Holznagel/Sonntag, in: Rossnagel (Org.), Handbuch Datenschutzrecht, 2003, p. 678 e ss., com outras indicações.

44 Neste sentido, BVerfGE 65, 1, 43.

45 Sobre o princípio do consentimento informado, v. § 4 § 1, inciso 1 BDSG, assim como art. 2 Lithdsrl.

tais mecanismos, como aqueles para a proteção da personalidade através da configuração do sistema e da tecnologia[46]. Uma vez que a pessoa envolvida é apenas, de forma limitada, o senhor do sistema e do projeto tecnológico, a proteção efetiva da privacidade pressupõe que a pessoa em questão possa geralmente confiar que tais mecanismos de proteção, na medida em que existam, serão realmente eficazes. Proteção dos direitos fundamentais da personalidade como proteção da liberdade então também exige a proteção da confiança, que vai além da proteção da confiança na possibilidade de autodeterminação decisão sobre a medida em que os dados podem ser acessados. Também deve ser assegurada a própria proteção da confiança na confidencialidade e na integridade do sistema técnico de informação ao qual o titular do direito fundamental se dirige, sem que se espere que ele possa controlá-lo.

IV – ABORDAGENS PARA A PROTEÇÃO DE DIREITOS FUNDAMENTAIS

O direito fundamental à personalidade – complementado igualmente por meio de outras normas de proteção, como na Convenção Europeia de Direitos Humanos (por exemplo, art. 8º da EMRK) –, possibilita que a comunicação baseada em técnicas de informação seja protegida como exercício da liberdade baseada na confiança.

IV.1 Defesa e proteção

A proteção dos direitos fundamentais inclui a defesa contra intervenções (injustificadas) do Estado. No entanto, também se trata da garantia de proteção, seja por meio do cumprimento dos direitos subjetivos inerentes aos direitos fundamentais, e eventualmente pelos respectivos deveres de proteção[47], seja pela configuração das prescrições jurídico-objetivas dos direitos fundamentais[48]. As dimensões de proteção fora da proteção puramente defensiva dos direitos fundamentais[49] tornam-se mais centrais para as garantias dos direitos fundamentais, quanto mais as reais condições prévias para o exercício da liberdade pelos cidadãos têm que ser criadas e mantidas

46 Sobre a proteção do sistema e suas distintas facetas, v. Albers, in: Hoffmann-Riem/Schmidt-Aßmann/Vosskuhle (nota de rodapé 2), § 22 número de margem 102 e ss.

47 Sobre deveres de proteção em geral, v., por exemplo, BVerfGE 39, 1, 42; 46, 160, 164; 56, 54, 73; 115, 118, 152.

48 Neste sentido, por exemplo, Stögmüller CR 2008, 435 e ss. Cf., também, Hornung CR 2008, 299, 305; Kutscha NJW 2008, 1042, 1044; Sachs/Krings JuS 2008, 486. 45.

49 Ver as indicações nas notas 43 e 45.

de um lado pelo Estado, e, por outro lado, também por privados, ou mesmo no curso de atos cooperativos entre Estado e privados, e que eventualmente podem ser questionadas por eles[50]. Portanto, está se tornando cada vez mais significativo o fato de que o Tribunal Constitucional Federal da Alemanha tem se reportado há tempo e por diversas vezes à dimensão jurídico-objetiva da proteção dos direitos fundamentais[51]. No entanto, nas mais recentes decisões do Senado da Corte sobre a proteção contra intervenções em comunicação estruturada tecnologicamente e contra o acesso às informações correspondentes, as intervenções ou autorizações de intervenções por parte do Estado[52] têm permanecido em primeiro plano, uma vez que só elas foram objeto nos respectivos processos. No que diz respeito à ativação de outros, ou seja, também a partir das funções jurídico-objetivas dos direitos fundamentais, ela exige – na medida em que não se tornem significativas no curso da interpretação e aplicação das normas válidas – como regra formulações correspondentes por parte do legislador. Para este fim, não estão abertas apenas proibições e imposições, mas também outras formas de configuração, como regulamentos em relação à organização e ao procedimento, ou sobre configuração técnica.

IV.2 Normas relevantes de direitos fundamentais

Distintas normas já estão à disposição para a proteção dos direitos fundamentais, como a proteção do sigilo das comunicações (art. 10 da LF), a inviolabilidade do domicílio (art. 13 da LF), assim como as diversas dimensões complementares e muitas vezes centrais do direito fundamental à proteção da personalidade a partir do art. 2º, § 1º, c/c art. 1º, § 1º, da LF[53], complementado eventualmente também pelos arts. 12 e 14 da LF, e subsidiariamente a liberdade geral de ação do art. 2º, § 1º, da LF.

50 Cf., por todos, Schulze-Fielitz, in: Hoffmann-Riem/Schmidt-Aßmann/Vosskuhle, Grundlagen des Verwaltungsrechts, vol. 1, 2006, § 23, especialmente número de margem 64 e ss., 91 e ss.

51 Cf., por exemplo – em relação ao art. 5 § 1º, inciso 1 da LF – BVerfGE 7, 198, 205 e ss.; ao art. 10 da LF BVerfGE 106, 28, 37; ao art. 2 § 1º em sentido amplo 1 § 1º da LF BVerfGE 96, 56, 64; ao art. 2 § 1º e 14 § 1º da LF BVerfGE 84, 192, 194 f. 114, 73, 89 e ss. Cf. também a argumentação para a abertura de uma fonte de informação no âmbito da liberdade informacional (art. 5 § 1º da LF): BVerfGE 103, 44, 61. Ver, além disso, BVerfGE 49, 89, 140 e ss., ou por exemplo BVerfG JZ 2007, 576.

52 BVerfGE 107, 299, 313 e ss. Afirma que as medidas das empresas privadas – aqui uma empresa de telecomunicações – devem ser imputadas ao Estado se tiverem sido ordenadas pelo Estado e a empresa em causa não tiver margem para alternativas.

53 BVerfG (nota de rodapé 17), número de margem 166 e ss.

IV.2.a Sigilo das telecomunicações

O art. 10 da Lei Fundamental protege a transmissão incorpórea de informações por meio do tráfego das telecomunicações[54]. O ponto de partida da garantia constitucional é a ideia de proteger os direitos e liberdades do indivíduo decorrentes do processo de transmissão técnica e do envolvimento de um intermediário de comunicação – em regra, devido aos perigos que surgem como resultado do distanciamento físico[55]. Com esse objetivo determinado, o direito fundamental contém, em especial, um direito de defesa contra a divulgação do conteúdo e das circunstâncias detalhadas das telecomunicações pelo Estado; mas também inclui a tarefa ao Estado de fornecer proteção contra o acesso de terceiros privados ao conteúdo e às circunstâncias da comunicação. Além disso, existe proteção contra o Estado tornar acessível a si mesmo o conhecimento relevante relacionado à comunicação de pessoas privadas, por exemplo, autorizando o acesso aos dados de tráfego (anteriormente: "dados de conexão")[56] detidos por empresas de telecomunicações em processos de comunicação específicos, ou por meio da padronização de uma obrigação de retenção de dados, juntamente com direitos de acesso aos dados armazenados[57].

IV.2.b Proteção do domicílio

A proteção também pode ser garantida pelo direito fundamental especial previsto no art. 13 da Lei Fundamental[58], que protege a esfera espacial em que se desenvolve a vida privada, especialmente contra intromissões, isto é, incluindo a utilização de meios para a aquisição de imagens e impressões sobre eventos no domicílio[59]. A proteção assim garantida estende-se à coleta de informações tornada possível pela intromissão, bem como ao uso dos dados obtidos desta forma.

54 Cf. BVerfGE 67, 157, 172; 106. 28, 35 e ss.; 115, 166, 182. Acerca da amplitude desta proteção, ver em especial Bäcker, in: Brink/Rensen (nota de rodapé 32), sob o título II.

55 Em todo caso, a distância espacial não pode ser um elemento essencial na medida em que a possibilidade de acesso não se baseia nela, mas sim no uso das telecomunicações – independentemente da distância que os computadores utilizados para eles estão "localizados" uns dos outros.

56 Cf. BVerfGE 107, 299, 312 f.; 113, 348, 365.

57 Sobre o tema, cf. §§ 113a, b TKG e a decisão do BVerfG NVwZ 2008, 543.

58 Cf. BVerfGE 89, 1, 12; 103, 142, 105 e ss.

59 BVerfG (nota de rodapé 17), número de margem 193.

IV.2.c Direito fundamental à proteção da personalidade

Particularmente importante é o direito fundamental à proteção da personalidade nos termos do art. 2º, § 1º, combinado com o art. 1º, § 1º, da Lei Fundamental[60], sobre o qual o Tribunal Constitucional já decidiu no julgamento do censo no sentido de que as concretizações realizadas até agora não são conclusivas. A complementação do direito à autodeterminação informativa não estava associada à declaração segundo a qual uma concretização conclusiva tivesse sido realizada.

O direito fundamental[61] à proteção da própria imagem, o direito fundamental à proteção da própria palavra, o direito fundamental à proteção dos dados pessoais e o direito fundamental à proteção da esfera privada em perspectiva espacial e temática, bem como o direito fundamental à autodeterminação informativa, há muito foram reconhecidos como manifestações parciais deste direito fundamental (não explicitamente contidos no texto constitucional).

Em sua decisão sobre as buscas on-line[62], o Tribunal Constitucional Federal a eles adicionou, como outra expressão parcial do direito fundamental, a garantia da confidencialidade e da integridade dos sistemas próprios de tecnologia da informação[63], por vezes designado direito fundamental da TI[64].

A relação destas manifestações parciais do direito fundamental à proteção da personalidade nem sempre é fácil de esclarecer. Os direitos fundamentais à sua própria imagem e à sua palavra são destinados a elementos da proteção da personalidade que também são cobertos pelo direito fundamental à autodeterminação informativa, mas que também são

60 Cf. também a nota de rodapé 4.

61 A designação como "direito fundamental" (ver, por exemplo, BVerfG NJW 2008, 1793, 1794) deve ser preferida à anteriormente "lei" consuetudinária, uma vez que enfatiza o fundamento constitucional e permite uma distinção à "lei" do direito civil correspondente. Quaisquer outras consequências legais não estão associadas a isso, conforme indicado em Böckenförde JZ 2008, 925, 927 nota de rodapé 25.

62 BVerfG (nota de rodapé 17): Especificamente, a sentença se refere a pesquisas online; mas seu alcance constitucional vai muito além disso.

63 A redução do direito fundamental como "direito fundamental do computador" proposta na mídia é enganosa. Melhor – mas inadequado como termo jurídico – é o direito fundamental de TI, que Bäcker, por exemplo, usa em: Brink/Rensen (nota 32).

64 Bäcker, in: Brink/Rensen (nota de rodapé 32).

cobertos por outros direitos fundamentais (como no caso do art. 5º da Lei Fundamental). A proteção da privacidade inclui dados pessoais[65], mas claramente vai além da sua proteção, por exemplo, quando se refere à proteção do comportamento em uma situação protegida como esfera privada, ou como proteção contra exigências comportamentais nos respectivos espaços. A garantia de confidencialidade e integridade dos próprios sistemas de tecnologia da informação, que agora foi reconhecida pelo Tribunal Constitucional, sobrepõe-se às outras subcategorias, mas ganha seu significado especial através do foco na proteção do uso dos sistemas de tecnologia da informação para fins próprios relacionados à personalidade contra ameaças associadas.

A proteção da confidencialidade e da integridade dos próprios sistemas de informática pelos direitos fundamentais não foi concebida pelo Tribunal Constitucional como um novo direito fundamental[66], mas como uma manifestação do direito fundamental à proteção da personalidade. Isto é, como os outros direitos mencionados da proteção da personalidade, não é explicitamente abordada na parte dos direitos fundamentais da Lei Fundamental, mas é fundada nela. O direito à proteção se baseia nas mesmas premissas normativas que são o fundamento das concretizações das outras dimensões protetivas do direito de personalidade.

V – EM ESPECIAL: A GARANTIA JURÍDICO-FUNDAMENTAL DA CONFIDENCIALIDADE E DA INTEGRIDADE DOS SISTEMAS TÉCNICOS DE INFORMAÇÃO DE USO PRÓPRIO

A nova forma de proteção da personalidade encontrou ampla aprovação na mídia[67], e tem recebido críticas na literatura especializada, mas

65 Isso parece ser o discurso da "privacidade eletrônica" de Böckenförde JZ 2008, 925. A objeção a isso, no entanto, é que a proteção da privacidade sob os direitos fundamentais – espacial/temática – não é definida com base no meio com o qual a privacidade é configurada.

66 Uma parte da literatura ignora isso, por exemplo Lepsius, em: Roggan (nota 34), p. 21 e ss. Este artigo também empreende uma reconstrução da decisão, que é tão destacada de suas declarações e premissas que a classificação dogmática constitucional de Lepsius não pode sequer começar a convencer. Portanto, Böckenförde JZ 2008, 925, 928 nota 38 também rejeita este ponto com razão.

67 Cf., por todos, Prantl Süddeutsche Zeitung aos 28.02.2008, p. 4.

também concordâncias[68]. Os críticos[69] consideram que a nova concretização é dispensável, especialmente porque a proteção pretendida já é concedida pelo direito fundamental à autodeterminação informativa; eles também veem o risco de que a proteção da autodeterminação informativa seja minimizada[70]. A falta de uma estrutura dogmática e os riscos associados na delimitação do direito à autodeterminação informativa são também criticados[71]. Há receios igualmente acerca de um "direito fundamental a-pessoal orientado para a tecnologia"[72]. A seguir, será feita uma tentativa de reconstruir as importantes premissas para o novo do direito fundamental e, em particular, de mostrar que a necessidade de proteção vai além do que aquela até agora desenvolvida segundo a jurisprudência com o direito fundamental à autodeterminação informativa.

V.1 O ponto de partida

Nas manifestações feitas até o momento pelo Tribunal Constitucional sobre a proteção do direito fundamental à autodeterminação informativa, foi afirmado, em particular, que tal direito oferece a seus titulares proteção contra coleta, armazenamento, uso e divulgação ilimitados de dados individualizados ou individualizáveis relacionados a eles[73]. Em parte, também

68 Em particular, a construção e a forma de raciocínio são atacadas em detalhes, mas não a dimensão de proteção pretendida. Da literatura bastante crítica, ver, por exemplo, Britz DÖV 2008, 411 e ss.; Sachs/Krings JuS 2008, 482 e ss.; Eifert NVwZ 2008, 521 e ss.; Lepsius, em: Roggan (nota 60), p. 21 e ss.; Bull, em: Anuário de Segurança Pública 2008/2009, p. 317 e ss.; V. MMR 2008, 365 e ss. Cf. também o acompanhamento nos nºs 16 e 25, bem como as contribuições em Roggan (nota 34). Em princípio e de acordo com muitos detalhes, por exemplo Hornung CR 2008, 299 e ss.; Hirsch NJW 2008 com referência a NJOZ 2008, 2902; Lorenz StRR 2008, 140 e ss.; Stögmüller CR 2008, 435 e ss.; Jäger Juris-itr 12/2008; Petri DUD 2008, 443; Bäcker, em: Brink/Rensen (nota 32); Böckenförde JZ 2008, 925 e segs.; Michael/Morlok, Grundrechte, 2008, número de margem 427 e ss.

69 Em particular, a visão subordinada do tribunal é criticada de que o direito à autodeterminação informacional só se aplica a "requisitos de comunicação individual ou dados armazenados" ou dados com "referência seletiva a uma determinada área da vida" (a este respeito, a referência é feita em particular para as formulações do Tribunal Constitucional [nota 17], número de margem 201 e ss.). No entanto, as declarações do Tribunal Constitucional são mal interpretadas se forem entendidas como as declarações finais sobre o âmbito da proteção do direito fundamental à autodeterminação informativa. Como o contexto das explicações mostra, deve, antes de tudo, ficar claro que a complexidade da necessidade de proteção no que diz respeito aos sistemas de tecnologia da informação ainda não foi adequadamente acolhida pela jurisprudência anterior. A jurisprudência e, em grande parte, também a literatura tratam de precauções contra medidas concretas de coleta e uso de dados, mesmo quando fornecem instrumentos – tais como precauções para autoproteção, proteção por meio de tecnologia e configuração de sistemas – em um nível anterior. A dimensão da proteção independente da confiança no sistema de tecnologia da informação utilizado pelo próprio Tribunal Constitucional não entra em foco.

70 Assim, por exemplo: Britz DÖV 2008, 411, 413; Sachs/Krings JuS 2008, 481, 484; Volkmann DVBl. 2008, 591; Eifert NVwZ 2008, 521 e ss.

71 Cf. Kutscha NJW 2008, 1043; Lepsius, in: Roggan (nota de rodapé 34); v. também nota de rodapé 60.

72 V., sobre o tema, acima nota de rodapé 36.

73 Cf. BVerfGE 65, 1, 43; 67, 100, 143; 84, 239, 279; 103, 21, 33; 115, 166, 190; 115, 320, 341 e ss.

foi afirmado (de forma transversal e, portanto, sem mais especificações e sem poder de delimitação jurídico-dogmática) no sentido de que devem ser levadas em conta as ameaças e violações da personalidade que surgem para o indivíduo, especialmente sob as condições do moderno processamento de dados, "a partir de medidas relacionadas à informação"[74]. As decisões sobre este direito fundamental tomadas pelo Tribunal Constitucional até o momento referem-se a ameaças causadas por coleta de dados, independentemente de ser seletiva ou contínua, ou se são realizadas em casos individuais ou em escala de massa.

As proteções que são possíveis, ou mesmo requeridas, não se limitam, no entanto, a medidas diretamente relacionadas ao processo de coleta de dados e subsequente armazenagem, uso, processamento ou divulgação, mas também se estendem aos pressupostos (organizacionais, procedimentais, sistêmicos, dentre outros) para garantir que tal coleta e subsequentes medidas atendam às exigências ou, conforme o caso, para que sejam encerradas. Aqui se torna claro que a proteção da autodeterminação informativa já começa no nível da ameaça aos direitos fundamentais e pode, portanto, ser alcançada por meio de medidas para reduzir tais ameaças. Mesmo quando as medidas de proteção – tais como medidas de proteção de dados do sistema[75] – estão previstas na coleta de dados, elas são medidas para evitar prejuízos aos dados – elas ocorrem especialmente na forma de controle e design do contexto, mas não na proteção da confiança no funcionamento do sistema de tecnologia da informação em si. Em outras palavras: a proteção de dados por meio do design do sistema não é idêntica à proteção do sistema de tecnologia da informação (independentemente de as disposições jurídicas para design do sistema serem implementadas nele) contra o acesso ao próprio sistema e o acesso subsequente aos dados.

Se forem formulados requisitos especiais para esta nova dimensão da proteção relacionada ao sistema, isto não implica, ao mesmo tempo, a crítica de parte da literatura referente à "minimização" do direito à autodeterminação informativa[76]: seu objetivo de proteção e seu nível de proteção permanecem inalterados. No entanto, seu âmbito de aplicação não se estende a outras dimensões de proteção que até o momento não foram suficientemente cobertas pelo direito fundamental à autodeterminação informativa;

74 Desta forma formulado pelo Tribunal Constitucional, NJW 2008, 1505, 1506 (sobre registros de placas).
75 Ver acima na nota de rodapé 44.
76 Ver acima na nota de rodapé 67.

em vez disso, essa proteção está ancorada em um novo direito fundamental (até este ponto especial) e implementada por meio de requisitos regulamentares mais rigorosos. Uma redução na proteção do direito fundamental da personalidade, em geral, de todo modo não se verifica.

V.2 Novas dimensões da necessidade protetiva

Na decisão sobre as buscas on-line, o tribunal partiu da premissa de que a proteção até então concebida para o direito fundamental à autodeterminação informativa não era suficiente para proteger a confiança importante para a proteção da personalidade na funcionalidade dos sistemas utilizados para comunicação. A proteção (somente) contra a coleta e posterior utilização de dados pessoais não é suficiente se não incluir também a proteção contra o acesso ao seu próprio sistema de tecnologia da informação, que é utilizado para o desenvolvimento comunicativo, cujo funcionamento sem perturbações é regularmente confiado, e cuja infiltração ou mesmo manipulação apresenta perigos para a proteção dos direitos de personalidade, que não são evitados pela proteção dos dados coletados[77]. Assim, a infiltração de um sistema complexo de tecnologia da informação com a possibilidade de manipulação de seu funcionamento ou de instalação de software para alterar os dados processados pelo sistema e os processos de comunicação transmitidos pelo sistema resultam em si em fontes de perigo, cujo surgimento também apresenta riscos para os dados disponíveis na tecnologia da informação. A defesa efetiva contra tais ameaças à personalidade requer um (pré)deslocamento da proteção para a infraestrutura utilizada, que deve assegurar a possibilidade de autodeterminação com os dados, bem como a liberdade e a integridade da comunicação transmitida através das infraestruturas. A necessidade de proteção contra tais infiltrações já existe antes

77 Que os perigos associados à violação da confidencialidade e integridade dos sistemas de tecnologia da informação sejam levados em consideração exclusivamente por meio da "proteção dos dados ao mesmo tempo suficientemente" – como B. Eifert NVwZ 2008, 522 assume – não é reconhecível. A proteção da confiança no tipo de desempenho deste sistema de tecnologia da informação não pode ser alcançada desta forma. Por exemplo, violações de integridade – como a manipulação do software com efeitos relacionados à proteção pessoal – podem tornar praticamente impossível a proteção de dados individuais. Além disso, a proteção que é (apenas) implementada como proteção dos dados coletados teria que se basear na qualidade desses dados, sem poder ser influenciada de forma independente em seu tipo e intensidade pela forma como foram obtidos. Deve-se admitir, no entanto, que o Tribunal Constitucional Federal, ao determinar a necessidade de proteção, em particular a determinação do nível de intervenção, também fez uso de circunstâncias que não estão relacionadas com a qualidade dos dados em causa, como a propagação ou o número de intervenções.

que certos dados possam ser acessados[78], e continua quando tal intervenção ocorre.

Diferentes níveis de perigo vêm à tona. Alguns dos perigos podem ser combatidos[79] pelo (já desenvolvido) direito fundamental à autodeterminação informativa, eventualmente após modificações; outros não poderiam, ou pelo menos não de tal forma que as especificidades da situação de risco no uso de seus próprios sistemas de tecnologia da informação são sejam suficientemente levadas em consideração.

V.2.a Amplitude da intervenção relativa à personalidade

O risco de que tais infiltrações possam facilitar mais a coleta de dados do que no passado poderia de fato ser combatido em muitos aspectos pelo direito fundamental à autodeterminação informativa – complementado também pelos arts. 10 e 13 da Lei Fundamental. Se a infiltração não só torna possível o acesso a processos de comunicação específicos ou a dados individuais, mas também a todos os outros dados "arquivados" no sistema de comunicação ou a dados que podem ser acessados através dele (por exemplo, dados do provedor), então podem ser capturadas uma infinidade e uma variedade de circunstâncias da vida e de características pessoais que antes dificilmente poderiam ser previstas em detalhe e possivelmente até tipificadas, o que só é possível em virtude da infiltração do sistema de tecnologia da informação. A amplitude "pessoal" de acesso ao sistema de tecnologia da informação aumenta o potencial de risco de intervenções posteriores em tecnologia da informação e reduz a possibilidade de contra medidas autodeterminadas. De todo modo, isso exigiria novos esforços dogmáticos caso tais riscos fossem combatidos unicamente pela extensão do alcance do direito fundamental à autodeterminação informativa.

V.2.b Sobre dados gerados pelo sistema

Em especial, é duvidoso se o direito fundamental à autodeterminação informativa é suficientemente eficaz contra o acesso aos dados relevantes para a personalidade gerados pelo sistema de tecnologia da informação – geralmente sem o conhecimento da pessoa envolvida. Acessos ao sistema de tecnologia da informação para fins de acesso a tais dados podem ser concebidos como uma intervenção no direito fundamental à autodeterminação

78 Neste sentido, Petri DUD 2008, 446.

79 V. as indicações acima na nota 66: vários autores qualificam este método como suficiente.

informativa e submetido ao seu programa de justificação. Os problemas surgem, no entanto, porque a possibilidade de proteção para a pessoa afetada – incluindo a possibilidade de proteção efetiva ex post – é, de fato, limitada. Não auxilia também, como é defendido em alguns casos, que seja referido à pessoa a possibilidade de autoproteção preventiva. Por exemplo, há certas possibilidades para o usuário de impedir tecnicamente a criação de dados individuais gerados no processo de comunicação – tais como cookies[80] – ou conjuntos de dados – tais como caches[81] –, mas apenas de forma limitada: eles sempre exigem tanto uma consciência especial do perigo quanto uma considerável experiência técnica, e em alguns casos – como no caso dos flash cookies – é extremamente difícil encontrá-los[82]. As qualificações correspondentes não podem ser adquiridas pelos usuários prontamente. Tampouco corresponde ao modelo constitucional de proteção da liberdade concedê-la apenas a uma pequena minoria de pessoas conscientes do perigo e com experiência técnica – tais como freaks, hackers, ou criminosos especializados em tais habilidades[83]. Também é importante observar que os cookies ou caches geralmente só podem ser desativados ao custo de uma perda não desprezível de funcionalidade: em muitos aspectos, eles também são "úteis" para a pessoa envolvida. Ela confia, em última análise, que pode utilizá-los sem preocupação.

Caso o usuário não queira impedir a geração e a coleta de dados, exige-se uma proteção efetiva da personalidade, de modo que o usuário tenha a confiança de que os dados assim obtidos não possam ser utilizados em contextos diferentes e, em particular, por terceiros não autorizados. Porém, por meio da infiltração dos sistemas técnicos de informática, eles podem ser utilizados por quem não está envolvido no processo de comunicação, sem que o interessado possa reconhecer e se proteger.

80 São dados armazenados em um computador usuário para transmitir certas informações a um computador servidor, especialmente no caso de visitas repetidas.

81 Um cache é uma memória de buffer rápida que contém cópias do conteúdo de outra memória (de fundo) e, portanto, acelera o acesso a ela. Portanto, os dados são armazenados em cache para acesso mais rápido a um meio mais rápido. A maioria dos navegadores da web cria esse cache no disco rígido na forma de arquivos temporários.

82 Os cookies em Flash – assim chamados em homenagem ao software flash player com o qual são criados – são consideravelmente mais difíceis de exibir e excluir em comparação com os cookies "normais". Com as configurações padrão do sistema operacional Microsoft Windows XP, por exemplo, eles não são iguais visíveis no disco rígido. Eles não podem ser localizados de forma alguma no navegador, pois são processados e armazenados independentemente do navegador. Pelo mesmo motivo, os dados armazenados podem ser atribuídos de forma clara ao respectivo usuário, mesmo quando usando navegadores diferentes no mesmo sistema e também em qualquer número de sessões do navegador.

83 Cf. BVerfG JZ 2007, 576: Proteção própria informacional precisa ser ao indivíduo possível e razoável.

V.2.c Criação de novas imagens da personalidade de nova profundidade e amplitude

Uma situação de risco especial, que não é coberta pela proteção tradicional dos direitos fundamentais sem extensões consideráveis, é criada pelo fato de que a infiltração supera fundamentalmente – ou seja, não apenas em casos individuais – obstáculos técnicos que, de outra forma, impediriam a espionagem ou vigilância. Se o obstáculo for superado, a barreira de proteção relacionada ao sistema, que de outra forma teria que ser superada repetidamente no caso de intervenções no direito à autodeterminação informativa e contra a qual pode haver opções de proteção legal, não está mais disponível. Mesmo que o ponto de infiltração esteja interessado apenas em determinados dados[84], é praticamente possível para ele também obter outros dados e acessar outros processos de comunicação. Por exemplo, a infiltração possibilita a obtenção de um banco de dados potencialmente grande, altamente informativo sobre várias facetas da personalidade. Percepções sobre partes essenciais do estilo de vida podem ser reveladas, bem como maneiras de criar perfis sociais, de interesse, comportamento e de comunicação diferenciados e, portanto, perfis de personalidade altamente significativos[85].

No entanto, o direito fundamental à autodeterminação informativa já protege contra a construção de imagens da personalidade por meio do uso de levantamentos de dados individuais[86]. No entanto, se a infiltração de sistemas de tecnologia da informação fundamentalmente remove o obstáculo técnico de acesso a todas essas informações, o registro de todos os dados acessíveis no sistema de tecnologia da informação por longos períodos de tempo cria oportunidades para o acúmulo e combinação de muitas informações de diferentes áreas da vida em uma profundidade e amplitude que não eram possíveis anteriormente com as intervenções[87]. Mesmo quando o direito fundamental à autodeterminação informativa em sua proteção contra imagens de personalidade é suficientemente ativado contra coleta de da-

84 Por exemplo, as autoridades de segurança estão regularmente interessadas apenas na transmissão seletiva direcionada de dados específicos que sejam relevantes para elas. Um Trojan instalado por você funciona melhor se apenas transmitir dados individuais, ou seja, precisamente os dados que são importantes para a tarefa oficial (por exemplo, nomes de parceiros de comunicação, conteúdo de e-mail salvo etc.). No entanto, o obstáculo é geralmente superado pela infiltração.

85 A redação em Böckenförde JZ 2008, 925, 928 é plástica: "É o serviço de mediação do sistema de tecnologia da informação que agrega dados pessoais individuais em um todo dinâmico que é acessível uma e outra vez e, portanto, expõe a pessoa em questão em sua vida pessoal em caso de acesso não autorizado".

86 BVerfGE 65, 1, 42 e ss.; 109, 279, 323; 112, 304, 319.

87 Michael/Morlok (nota de rodapé. 64), número de margem 429 falam sobre um "salto qualitativo".

dos específicos, a infiltração de sistemas de tecnologia da informação ainda envolve riscos de que imagens de personalidade de amplitude e densidade anteriormente desconhecidas sejam criadas e de que a pessoa interessada nem mesmo avalie o perigo potencial e, muitas vezes, sequer consiga se defender com eficácia: em qualquer caso, a lacuna de proteção relacionada ao sistema de tecnologia da informação não se deixa ser efetivamente eliminada no nível de proteção contra a coleta de dados específicos. A infiltração do sistema de tecnologia da informação colocou um "pé virtual na porta" da personalidade.

V.2.d Risco de falsificação de dados

A possibilidade de infiltração no sistema também está associada ao risco de falsificação (praticamente irreconhecível) dos dados individuais registrados e sua combinação com outros, o que, neste aspecto, pode levar a um perfil de personalidade falsificado. O afetado praticamente não pode mais se defender contra tais falsificações, uma vez ocorrida uma infiltração associada a tais possibilidades, que, em princípio, também pode ser utilizada por terceiros[88]. Isso não é, de forma alguma, apenas uma intensificação da intervenção contra a qual o direito fundamental à autodeterminação informativa protege em seu conteúdo[89], mas uma qualidade independente de ameaça[90]. As medidas de proteção necessárias devem começar pela proteção do próprio sistema de tecnologia da informação, ainda que essa proteção deva então ser estendida também aos dados coletados em decorrência da infiltração, no interesse de sua eficácia.

V.2.e Neutralização das possibilidades de autoproteção

A infiltração estatal e – se necessário – para espionagem – a manipulação do sistema de tecnologia da informação devem, em particular, levar ao fato de que a autoproteção aplicada e mesmo recomendada ao afetado como uma expressão da ideia de autodeterminação informativa – por exemplo, a criptografia e o uso de senhas – será fundamentalmente (não só no caso concreto) driblada. A proteção proporcionada pelo direito fundamen-

88 O fato de que as autoridades de segurança que executam a infiltração não deveriam estar interessadas em tais falsificações é apenas mencionado para fins de completude.

89 Por exemplo, Eifert NVwZ 2008, 521: "interferência particularmente grave" na área de proteção do direito fundamental à autodeterminação informativa.

90 Se fosse apenas sobre o problema de "violações aditivas aos direitos fundamentais", poderia, no entanto, ser tratado também no nível da justificação. Ver, por exemplo, BVerfGE 112, 304, 319 e seguintes. No entanto, isso é mais do que apenas uma adição.

tal à autodeterminação informativa será comprometida em suas premissas básicas.

Isso porque a possibilidade de autoproteção tem sido considerada até agora um elemento essencial da participação autodeterminada na comunicação, que também é levada em consideração no direito atual da proteção de dados. A possibilidade de autodeterminação sobre os dados disponíveis foi considerada, por exemplo, pelo Segundo Senado do Tribunal Constitucional como ensejo para a tese segundo a qual a proteção conferida pelo art. 10 da LF para dados no âmbito de controle da pessoa em causa não se aplica, uma vez que ela tem possibilidades de autoproteção[91]. Se a referência do Segundo Senado à possibilidade de autoproteção é praticável e, portanto, realmente funciona, pode ser questionada. No entanto, é viável a reflexão de que os dados armazenados após a conclusão do processo de comunicação não diferem mais daqueles contidos nos arquivos criados pelo próprio usuário. Se, após a conclusão de um processo de comunicação, forem acessados dados de comunicação armazenados no domínio do destinatário, então se materializa não um risco específico de comunicação, mas um risco geral de tecnologia da informação[92].

O alto padrão da possibilidade de autoproteção determinado pelo direito básico à autodeterminação informacional não é desvalorizado pelo fato de muitos cidadãos tratarem seus dados de forma descuidada ou desconsiderarem as possibilidades de autoproteção. A necessidade de proteção dos direitos fundamentais não deixa de existir porque os cidadãos individuais não a sentem ou não podem realizá-la; à autodeterminação pertence a capacidade de decidir o quanto alguém deseja se proteger. Quem quiser prescindir da proteção também faz uso do direito à liberdade. Se, no entanto, ele não puder mais estimar a necessidade de proteção ou se a possibilidade de proteção nem mesmo existir, sua disposição para a proteção autodeterminada não importa mais e a recusa fundamental de proteção em nenhuma circunstância pode ser justificada com referência ao descuido de muitos cidadãos no tratamento dos seus dados (específicos). A possibilidade de proteção é, no entanto, negada aos cidadãos pela infiltração dos sistemas de tecnologia da informação. Isto se aplica mesmo que não ocorra em segredo, desde que o afetado não possa avaliar as consequências da

91 BVerfGE 115, 166, 185 e ss.

92 Desta forma elaborado por Bäcker, in: Brink/Rensen (nota de rodapé. 32), sob o título II 2c.

infiltração e da manipulação associada ou seja praticamente incapaz de as contrariar.

V.2.f Possibilidade de acesso de terceiros

Em particular, existe a necessidade (que já foi apontada por vezes aqui) de proteção contra o risco de terceiros (privados) se aproveitarem da infiltração do sistema de tecnologia da informação por parte das autoridades estatais e, por exemplo, se utilizem do software infiltrado para espionar o sistema ou manipulá-lo – isto é, de rededicar a infiltração a fins próprios como uma espécie de ovo do cuco posto pelo Estado, sem que a pessoa afetada suspeite disso e seja capaz de se proteger com eficácia. A proteção dos direitos fundamentais contra intervenções do Estado – aqui a infiltração – é constitucionalmente mais abrangente, mais fácil e, acima de tudo, mais eficazmente executável do que a proteção contra particulares no curso da eficácia horizontal indireta dos direitos fundamentais. No caso de uso estatal do software infiltrado pelo Estado ou do software ou hardware manipulado, há também a perspectiva de que o Estado observe as restrições constitucionais à sua autorização para intervir; no caso de acesso (ilegal) de terceiros viabilizado por "adiantamentos" estatais, essa perspectiva de proteção não se aplica, já que terceiros não se submetem a tais obrigações constitucionais e, dada a ilegalidade de seu comportamento, dificilmente poderiam ser efetivamente submetidos.

V.2.g Grande variação de pessoas envolvidas

A infiltração e o que é tornado possível por ela, se necessária a vigilância de longo prazo que abrange uma ampla variedade de atos de comunicação e os vincula dinamicamente, não se limita aos seus destinatários como afetados, mas inclui um grupo de terceiros que não pode ser esquecido de antemão como parceiros de comunicação da pessoa envolvida. Este também é o caso com outros tipos de acesso à comunicação – por exemplo, por meio de escuta telefônica ou observação policial. Na medida em que todos os tipos de dados relativos a terceiros são armazenados ou gerados no sistema de tecnologia da informação, a propagação pessoal que é possível aqui pode, no entanto, exceder qualitativamente aquela que está associada à intervenção direcionada em atos de comunicação específicos, como espionagem de certas conversas. Como resultado, terceiros podem ser afetados não apenas na medida em que isso seja "inevitável em casos individuais", mas potencialmente em princípio e – presumivelmente com

frequência – sem qualquer limitação prévia e – é claro – sem serem capazes de se defender "autodeterminadamente".

V.3 Esclarecimento da peculiaridade da situação de risco e da respectiva proteção de direitos fundamentais pelo Tribunal Constitucional Federal da Alemanha

V.3.a Diferenciação ao nível do âmbito de proteção

É verdade que uma tentativa poderia ser feita para lidar com algumas das preocupações que acabamos de listar por meio de uma expansão posterior do direito fundamental à autodeterminação informativa. Teria então de ser desenvolvido em um baluarte que pode ser usado de forma abrangente contra a habilitação e implementação do acesso estatal não apenas aos dados e processos de comunicação de todos os tipos, mas também às infraestruturas de comunicação utilizadas (software e hardware) e também contra os correspondentes acessos de particulares. No interesse da capacidade de gestão dogmática, seria necessária uma maior diferenciação do amplo âmbito de proteção com uma dogmática dos limites correspondentemente coordenada, com a elaboração de limites especiais (em regra especialmente mais altos) para a infiltração e manipulação de sistemas de tecnologia da informação que colocam em perigo a proteção pessoal e a coleta e o processamento de dados que isso permite. Em contraste, parecia constitucionalmente preferível para o Tribunal Constitucional diferenciar ainda mais o direito fundamental geral à proteção da personalidade e a proteção da integridade e confidencialidade dos sistemas de tecnologia da informação usados e antes da coleta e uso dos dados obtidos como resultado do infiltração em uma forma "especial", baseada no sistema de tecnologia de informação, do direito fundamental geral, que não depende de ficções de proteção autodeterminada da personalidade, mas antes coloca, em primeiro plano, a necessidade de proteção da confiança. Isso torna mais fácil observar a nova qualidade da ameaça e a necessidade de proteção baseada na confiança do sistema no nível do âmbito de proteção e a necessidade de reconhecer requisitos especiais para limites e de desenvolver medidas de proteção voltadas para a ameaça.

A abordagem do tribunal também pode ser interpretada como uma reação ao fato de que as dimensões da ameaça à confiança nas infraestruturas de comunicação e as necessidades de proteção correspondentes até então só foram abordadas em uma extensão limitada – se é que o fizeram – e que não existem conceitos, aprofundadamente discutidos ou reconhecidos na

jurisprudência e na literatura, sobre como a proteção da confidencialidade e da integridade dos sistemas de tecnologia da informação de uso próprio pode ser embutida no direito fundamental à autodeterminação informativa sem inconsistências e lacunas. Tendo em vista a falta de trabalhos anteriores na literatura, é surpreendente que a maioria dos autores que analisam a nova decisão alega, sem maiores diferenciações, que a proteção poderia ter sido realizada unicamente pelo direito fundamental à autodeterminação informativa. Isso é tanto mais surpreendente quanto o fato de que, antes da decisão na literatura e nos escritos submetidos ao tribunal, foram feitas tentativas para satisfazer a necessidade de proteção em particular por meio do art. 13[93] da Lei Fundamental – ou mesmo também do art. 10 da LF.

V.3.b Reação à especial qualidade do risco

Por outro lado, a designação explícita e a ênfase da proteção dos direitos fundamentais de confidencialidade e integridade dos próprios sistemas de TI, defendida pelo Tribunal Constitucional, deixa claro que qualitativamente há uma situação de risco especial e que as precauções de proteção correspondentes devem estar em vigor. Especificar o âmbito de proteção tem a vantagem, dentre outras, de que o teste de proporcionalidade no sentido amplo pode ser orientado de forma mais precisa. O potencial de risco particular é identificado destacando a expressão de direito fundamental particular de uma maneira tipificadora e a demanda por proteção tipificadora é feita. Como resultado, a proteção não depende apenas de ponderações ad hoc no contexto de testes de proporcionalidade. Ponderações relacionadas ao caso, no entanto, ainda podem ser necessárias para o ajuste estrito em casos individuais.

V.3.c Delimitação com relação ao direito à autodeterminação informativa

Um problema, entretanto, é a demarcação entre o direito à autodeterminação informativa e a proteção da integridade e da confidencialidade dos próprios sistemas de TI tratados aqui. O princípio básico é: contra a coleta de dados (e processamento posterior de dados)[94] sem a infiltração de sistemas de tecnologia da informação e contra a criação das autorizações correspondentes, o direito fundamental à autodeterminação informativa

93 Cf. os argumentos em Böckenförde JZ 2008, 925, 926 nota de rodapé 10.

94 No entanto, a utilização por terceiros, ou seja, após o repasse, é, de acordo com princípios gerais, apenas admissível se os pré-requisitos que justificam tal intervenção também forem cumpridos por esses órgãos.

continua a proteger[95-96]. Se, no entanto, um complexo sistema de informática for infiltrado, espionado e possivelmente manipulado para realizar a coleta de dados, entra em vigor a nova dimensão da proteção aos direitos fundamentais[97]. Este direito fundamental de proteger a confidencialidade e integridade do sistema de tecnologia da informação não só afeta a infiltração (e possivelmente a manipulação) como tal, mas também se estende à coleta e uso dos dados e informações que são (apenas) obtidos como resultado da infiltração[98]: os obstáculos respectivamente aumentados da proteção da personalidade estendem-se ao tratamento dos dados relacionados à personalidade acessíveis por meio da infiltração.

V.3.d Necessidade de outras concretizações

Os contornos da nova especificação dos direitos fundamentais não puderam ser trabalhados em todos os aspectos pelo Tribunal Constitucional, que se ocupou de um litígio específico e na medida em que fez referência ao objeto desse litígio. Como resultado, existe ainda uma necessidade considerável de especificações adicionais, também no que diz respeito ao objeto da proteção, em particular o conceito (relacionado com a personalidade) dos próprios (melhor: de uso próprio)[99] "sistemas de tecnologia da informação"[100]. Também não foi ainda definitivamente esclarecido de que forma deve ser salvaguardada a proteção dos sistemas informáticos contra intervenções não encobertas, que o tribunal também mencionou expres-

95 É errado interpretar a sentença introdutória do Tribunal Constitucional Federal (nota de rodapé 2), números de margem 166 e 201, no sentido de que a nova dimensão da proteção é "subsidiária" ao direito à autodeterminação informacional, por exemplo Petri DUD 2008, 444. O Tribunal Constitucional Federal afirma, em vez disso, que a nova forma de direitos fundamentais se aplica quando uma lacuna na proteção for diagnosticada.

96 A propósito, no que diz respeito a qualquer concorrência remanescente, aplica-se o princípio geral de que os limites dos direitos fundamentais devem ser derivados da expressão do direito de personalidade que protege contra o perigo maior e, portanto, impõe requisitos mais rígidos. Para informações gerais sobre tais regras de competição, ver Jarass/Pieroth, GG, 9ª edição 2007, observações preliminares antes do art. 1, número margem 18, com comentários adicionais.

97 Caso a autorização legal permita que outros órgãos se aproveitem da infiltração ou dos dados obtidos por meio dela, os elevados requisitos para interferir em seu próprio sistema de tecnologia da informação também teriam que ser atendidos por eles.

98 Esta extensão da proteção que foi "obtido" através da intervenção dos direitos fundamentais não é incomum. O art. 13 da LF não protege apenas contra a intrusão no domicílio, mas também as informações ou objetos obtidos por meio da intrusão, ver BVerfGE 109, 279, 374 com referência a BVerfGE 100, 313, 360 (este último no art. 10 da LF). Sobre os paralelos entre a nova garantia dos direitos fundamentais e o art. 13 da Lei Fundamental, ver Bäcker, em: Brink/Rensen (nota de rodapé 32), ponto III 1; Pieroth/Schlink, Grundrechte, 24ª edição 2008, número de margem 377c.

99 Essa alternativa linguística apenas evita ecos inadequados do direito das coisas. Ver também Bäcker, em: Brink/Rensen (nota 32), título III 2a. Há também (mesmo que apenas temporário) uso pessoal ao usar o computador no cibercafé.

100 V. também os requisitos acima III.

samente, mas não foi desenvolvida[101]. É também necessário esclarecer o âmbito da proteção contra particulares. A formulação constitucional da dimensão da proteção dos direitos fundamentais como "garantia" deixa claro, no entanto, que o Estado também tem a responsabilidade de garantir que a integridade e a confidencialidade dos sistemas de tecnologia da informação sejam protegidas na medida em que sejam ameaçados de outras formas que não pela intervenção estatal. No entanto, ele tem uma ampla margem de manobra criativa para a execução dos mandatos regulatórios legais e objetivos correspondentes.

A garantia de direito fundamental também protege contra intervenções com fins repressivos. No entanto, são necessários mais esclarecimentos sob quais condições isso pode ser possível[102]. Ao fazê-lo, o peso dos interesses jurídicos, cuja proteção de que efetivamente serve a norma penal eventualmente violada no caso concreto, deverá ser apurado de forma análoga às medidas preventivas.

No entanto, houve uma necessidade de especificações adicionais também ao formular o direito fundamental à autodeterminação informativa há um quarto de século. Lá, também, a nova perspectiva representou um desafio à dogmática jurídica, à legislação e à jurisdição. Então, agora ele está de volta.

VI – LIMITES AOS DIREITOS FUNDAMENTAIS

O direito fundamental à garantia da integridade e confidencialidade dos sistemas de tecnologia da informação de uso próprio não é protegido sem limitações. Tendo em vista o potencial de risco particular, o teste de proporcionalidade especialmente em regra (mas dependendo da intensidade da intervenção[103]) leva a um grande obstáculo para as intervenções. O dever de proteção do Estado, ancorado no direito objetivo, também é acionado para tomar medidas contra os perigos apresentados por particulares[104].

101 Cf., sobre o tema em mais detalhes, Böckenförde JZ 2008, 925, 931; Bäcker, in: Brink/Rensen (nota de rodapé 32), título III.

102 Cf., também die, as ponderações de Kühne, in: Roggan (nota de rodapé 34), p. 85 e ss.

103 O BVerfG não teve que decidir em que medida as intervenções de menor alcance do que as pesquisas online poderiam ser permitidas em condições menos estritas. Ver também Bäcker, em: Brink/Rensen (nota 32), sob III, 3.

104 Ver as notas acima 44, 45. Com o termo "garantia" da confidencialidade e integridade dos sistemas de tecnologia da informação, o tribunal esclarece a existência de um mandato ao estado para proteção em todas as áreas da vida (ver também Petri DUD 2008 446 e ss.), mas sem elaborá-lo com mais detalhes.

VI.1 Requisitos jurídicos materiais e processuais

O Tribunal Constitucional formulou requisitos para autorizações legais formuladas na área da prevenção de perigos, que dizem respeito à infiltração e manipulação do sistema informático utilizado pelo próprio usuário, mas também dizem respeito à coleta e utilização dos dados e informações obtidos nesta base.

Os requisitos constitucionais das restrições incluem, em primeiro lugar, o cumprimento do requisito da especificidade e da clareza das normas de autorização, que desde sempre é derivado do mandamento do Estado de Direito[105].

Os requisitos para a classificação do bem jurídico protegido também são importantes. Um bem jurídico suficientemente (predominante) para justificar uma busca online[106] inclui a vida, a integridade e a liberdade de pessoas ou bens do público em geral, cuja ameaça afeta os fundamentos ou a existência do Estado ou os fundamentos da existência das pessoas[107]. Um exemplo do último são os ataques a instituições públicas de segurança social, como as represas.

Existem também requisitos constitucionais para o tipo e intensidade do risco e, portanto, também para o grau de probabilidade e a base factual do prognóstico do risco[108]. Em particular, a exigência de uma probabilidade suficiente de ocorrência não pode ser dispensada e as suposições e conclusões devem ter um ponto de partida concreto de fato. Os fatos devem, por um lado, permitir concluir que pelo menos o seu tipo se concretize e seja temporalmente previsível e, por outro lado, que estarão envolvidas certas pessoas cuja identidade é conhecida pelo menos o suficiente para que a medida de vigilância possa ser usada especificamente contra elas e em grande parte limitada a elas[109].

Além disso, as garantias processuais são importantes[110], em particular um controle por uma autoridade independente, que é fundamentalmente

105 Cf. BVerfG (nota de rodapé 17), número de margem 208 e ss., com outras indicações.

106 Requisitos mais baixos podem ser suficientes, por exemplo, para a avaliação offline do disco rígido de um computador confiscado.

107 BVerfG (nota de rodapé 17), número de margem 247.

108 BVerfG (nota de rodapé 17), número de margem 242 e ss., 249 e ss.

109 BVerfG (nota de rodapé 17), número de margem 251.

110 BVerfG (nota de rodapé 17), número de margem 257 e ss.

necessário defronte a infiltração. O acesso secreto aos sistemas de tecnologia da informação, que podem ser avaliados como particularmente importantes, deve fundamentalmente estar sujeito a uma reserva de ordem judicial. Exceto em casos urgentes, outro órgão só pode ser considerado neste caso se oferecer a mesma garantia de independência e neutralidade de um juiz – uma garantia difícil de ser estruturada. As razões da legalidade das medidas de vigilância devem ser registradas por escrito.

VI.2 Núcleo da vida privada

Finalmente, precauções para proteger o núcleo essencial da vida privada são indispensáveis. Os sistemas de tecnologia da informação usados exclusivamente para a comunicação relevante para a área central não devem ser infiltrados. No entanto, isso geralmente não é previsível com antecedência. Nesse sentido, a proteção só pode ser totalmente eficaz quando os dados são coletados devido à infiltração do sistema de tecnologia da informação.

A coleta de dados relevantes para a área central deve ser, em princípio, evitada. A proteção só pode ser adiada para o segundo nível, designadamente a avaliação, se a relevância do núcleo central dos dados coletados não puder ser esclarecida antes ou durante a coleta de dados, mas houver indícios de uma suposta ameaça de perigo de um bem protegido de extrema importância. Ao fazê-lo, no entanto, regras procedimentais adequadas devem assegurar que a intensidade da violação do núcleo essencial e seus efeitos sobre a personalidade e o desenvolvimento da pessoa em causa permaneçam tão baixos quanto possível[111].

A proteção pela não coleta continua sendo a prioridade. Assim, o tribunal formula a exigência de abster-se de coletar dados se houver indícios de que o núcleo essencial está "afetado". O núcleo essencial da vida pessoal é protegido como tal. Não se trata (apenas) da proteção de uma determinada declaração que deve ser avaliada isoladamente, a qual, pelo seu conteúdo, não deve ser acessível ao Estado por razões de dignidade humana. Em vez disso, a proteção do núcleo essencial visa proteger aquela parte do desenvolvimento pessoal privado que, em prol da dignidade humana, deve ser mantida livre do conhecimento do Estado. Mesmo no núcleo essencial da vida privada, no entanto, o íntimo e o banal, o pessoal e o menos significati-

111 BVerfG (nota de rodapé 17), número de margem 281 e ss.

vo, normalmente se misturam. Essa situação de confusão comunicativa também é protegida, mesmo antes de se fazer o levantamento da comunicação, não apenas ao nível da avaliação. Lá, a proteção só poderia ser concedida dividindo o processo de comunicação em – visto isoladamente – conteúdo absolutamente protegido e conteúdo apenas relativamente protegido.

Se a proteção fosse oferecida apenas desta forma, o Estado, em princípio, não estaria impedido de se infiltrar no sistema de tecnologia da informação e, em primeiro lugar, registrar todo o conteúdo para depois remover as declarações individuais como absolutamente protegidas. Isso não faria justiça à ideia básica da proteção do núcleo essencial: a dignidade humana exige que o Estado se abstenha de monitorar uma situação em que haja indícios de que bem de proteção da mais alta prioridade seja afetado pela medida. Esse "contato" geralmente ocorre quando o sujeito está ciente dele. Uma renúncia à proteção no nível da coleta deve, portanto, permanecer uma exceção[112], para a qual há um ensejo, por exemplo, se o núcleo essencial for afetado inesperadamente[113] ou se houver indicações de que a comunicação serve para atender ou planejar atos criminosos específicos[114], ou porque conteúdos íntimos ou outros que precisam de proteção servem apenas como uma camuflagem para que as ações que dão origem a perigos sejam acordadas ou discutidas em mais detalhes[115]. Somente se não for suficientemente previsível qual conteúdo os dados coletados terão, ou se as dificuldades de tecnologia da informação ou de técnica de investigação impedirem a análise do conteúdo dos dados – por exemplo, no caso de documentos em língua estrangeira ou conversas –, permite-se, no que diz respeito aos bens extremamente importantes, mesmo após uma infiltração no sistema de tecnologia da informação, primeiro fazer um levantamento e deixar a proteção constitucional para o nível de avaliação (conceito de proteção em duas etapas).

O requisito de proteção do núcleo essencial não é cumprido pelo fato de que a coleta (somente) é evitada se "apenas" as descobertas relevantes ao núcleo central forem afetadas, conforme previsto na Seção 100a (4) do Código de Processo Penal e em outras normas. É extremamente raro que o conteúdo relevante para o núcleo essencial seja comunicado "sozinho" na

112 BVerfG (nota de rodapé 17), número de margem 281.
113 BVerfGE 109, 297, 318.
114 BVerfGE 113, 348, 391.
115 BVerfG (nota de rodapé 17), número de margem 281.

vida prática em algum ponto; o que não ficará aparente de antemão. Limitar a proteção a esse tipo de uso prejudicaria a proteção do núcleo central de dois estágios. Mesmo em uma conversa confidencial entre cônjuges, em que o conteúdo relacionado à área central é o assunto, também haverá outros conteúdos banais, por exemplo, declarações sobre o comportamento de terceiros ou acontecimentos de outro tipo: não atende aos requisitos constitucionais só por esse motivo permitir o monitoramento e a gravação e negar a proteção do núcleo central ao nível da coleta, adiando-a para o nível de aplicação[116].

VII – CONCORRÊNCIA COM OUTRAS NORMAS DE DIREITOS FUNDAMENTAIS

O direito fundamental à garantia da confidencialidade e integridade dos sistemas de tecnologia da informação pode concorrer com outras normas de direitos fundamentais, como os arts. 10 e 13 da Lei Fundamental.

VII.1 Intervenção no âmbito do domicílio

O âmbito de proteção do art. 13 da LF é afetado se houver uma intervenção no domicílio (ver IV 2b acima). Além disso, há obstáculos especiais para as intervenções, em particular as do § 4. No entanto, é duvidoso que o art. 13 da LF possa garantir a proteção de forma abrangente e adequadamente diferenciada, em particular se a proteção espacial do art. 13 da LF possa causar problemas específicos de infiltrações e alterações em sistemas de tecnologia da informação registrados: art. 13 da LF concede proteção espacial e proteção comportamental referente ao espaço, mas não proteção funcional relacionada à infraestrutura de comunicação pessoal. Aliás, a proteção sob o art. 13 da LF só entraria em consideração se o hardware infiltrado estivesse em um domicílio, situação que nem sempre ocorre em especial com notebooks e smartphones, exemplificativamente.

Deve-se acrescentar, no entanto, que o art. 13 da Lei Fundamental assegura aspectos importantes de proteção, como proteção contra a medição de emissões eletromagnéticas para capturar palavras de código, contra

116 Já no BVerfGE 113, 348, 391 e seguintes, diz – porém com relação ao art. 10, § 1, da LF, que basicamente concede uma proteção mais fraca do que o direito fundamental recentemente concretizado – "No caso concreto há indicações reais para a suposição de que uma vigilância de telecomunicações. Se for registrado conteúdo pertencente a esse núcleo essencial, ele não pode ser justificado e deve ser omitido". A palavra "captura" é usada aqui, mas não com base no fato de que a comunicação "sozinha" contém conteúdo na área central da vida privada. Essa afirmação não foi corrigida pela sentença da busca online.

intrusão na residência, por exemplo, para fins de manipulação do dispositivo, ou contra o acionamento de câmeras e microfones em computadores para monitoramento de atividades no domicílio[117]. Tais medidas constituem também uma intervenção independente que carece de justificação no domínio da proteção do art. 13, § 1º, da Lei Fundamental quando servirem à implementação prática da infiltração de sistemas de tecnologia da informação, tal como uma "busca online". O domínio da proteção dos direitos fundamentais da confidencialidade e integridade dos sistemas de tecnologia da informação não deve ser mal interpretado como se fosse suprimir as garantias dos direitos fundamentais que são afetados paralelamente, de modo que as medidas seriam admissíveis em anexo, na medida em que servem para implementar uma intervenção em um sistema de tecnologia da informação e isso seja permissível como tal – em relação ao padrão de garantia de proteção da personalidade de que trata aqui. De um lado, tal "solução do anexo" não faria justiça ao estatuto do direito fundamental de inviolabilidade do domicílio, que é particularmente protegido pela constituição, por exemplo com uma reserva judicial garantida. De outro, ela também não pode convencer sistematicamente, uma vez que a infiltração de sistemas de tecnologia da informação é tecnicamente possível sem entrar no domicílio[118].

VII.2 Concorrência com o sigilo de telecomunicação especialmente com as TKU-fonte

Existe uma situação de concorrência entre o art. 10 da Lei Fundamental e o direito fundamental à garantia da confidencialidade e integridade dos sistemas de tecnologia da informação, particularmente no caso de monitoramento de telecomunicações na fonte (TKÜ-fonte). O monitoramento de telecomunicações na fonte é um processo de monitoramento que registra as telecomunicações de saída antes da criptografia ou as telecomunicações de entrada após a decifragem pelo destinatário. Enquanto o monitoramento de telecomunicações costumava ocorrer e ser bem-sucedido no período de transmissão na rede – especificamente no caminho de transmissão –, isso não é mais possível com a transmissão digitalizada e o uso de criptografia.

117 Ver – embora não para todos os exemplos mencionados acima – BVerfG (nota 17), parágrafo 193.

118 Sobre o lado técnico, consulte Buermeyer HRRS 2007, 154, 163 e ss. Böckenförde JZ 2008, 925, 933 nota de rodapé 95 enfatiza que o acesso online, sujeito aos princípios da proporcionalidade, pode superar a instalação de hardware na residência.

No entanto, ainda falta clareza sobre a segurança contra a escuta de tecnologias individuais de Voice-over-IP[119].

O monitoramento de telecomunicações na fonte pode levar a perigos que vão além do monitoramento das telecomunicações em andamento durante a transmissão da rede[120]. O Tribunal Constitucional assumiu que as situações de ameaça não podem ser combatidas de forma adequada pelo art. 10º, § 1, da Lei Fundamental, porque existem dados que, na sequência de uma infiltração, são coletados sem referência a telecomunicação em curso. Ao mesmo tempo, afirmou que o art. 10 da LF é o único padrão de teste, desde que apenas a telecomunicação em andamento seja abrangida. A ideia básica dessa declaração é que a circunstância técnica de se o monitoramento ocorreu durante a transmissão da rede ou no terminal pode não ter nenhum significado para a atribuição do art. 10 da LF se a intervenção se limitar à captura da comunicação corrente e, portanto, o potencial de risco específico para a confidencialidade e integridade de sistemas complexos de tecnologia da informação não é ativado. No entanto, o Senado acrescentou que essa restrição deve ser resguardada por precauções técnicas e também garantida em termos legais[121].

No entanto, é duvidoso que tais precauções técnicas sejam possíveis atualmente. Na audiência de 10 de outubro de 2007, diversos especialistas ouvidos pelo tribunal negaram; na literatura, existem também vozes afirmativas[122]. No entanto, já há dúvidas se é praticamente possível se infiltrar em um sistema de tecnologia da informação sem obter um mínimo de informações – por exemplo, sobre seus pontos fracos; o conhecimento de tais pontos fracos podem desencadear novas ameaças. Acima de tudo, há dúvidas se as arquiteturas atuais ou previsíveis de computador permitem tal limitação de acesso: uma vez que um software é executado em um sistema, ele pode basicamente ser usado universalmente.

---

119 O software "Skype", por exemplo, há muito é considerado à prova de bugs e um excelente exemplo da necessidade de telecomunicações de origem. Entretanto, há evidências crescentes de que existe uma "chave duplicada" que também pode ser usada pelas autoridades para o processo de criptografia secreta, de modo que a escuta secreta também seria possível sem uma fonte TKÜ; consulte http://www.heise.de/newsticker/meldung/113 281.

120 BVerfG (nota de rodapé 17), número de margem 188 e ss.

121 BVerfG (nota de rodapé 17), número de margem 190.

122 Em sentido afirmativo, por exemplo, Bär MMR 2008, 423, que considera a existência de um software especial que só se abre quando as chamadas são efetivamente feitas e não requer o acesso a quaisquer outros dados do computador. No entanto, esta é a redação do art. 20 l, § 2º, Cláusula 1, nº 2 do Projeto de Lei do BKA, BR-Drs. 404/08 AC 5 de junho de 2008.

Os requisitos para que uma medida afete apenas o âmbito de proteção do art. 10 da LF e, portanto, para que somente esta norma seja a norma de teste não são atendidos em nenhum caso se a vigilância das telecomunicações estiver dependente de uma infiltração no sistema de tecnologia da informação, o que causa e pode causar intervenções relevantes à personalidade. O mesmo se aplica se o risco de uma alteração técnica no sistema for criado por infiltração ou como resultado do seu uso por terceiros. Essas ameaças à proteção da privacidade não podem ser defendidas apenas com o art. 10 da LF.

O teste à luz do direito fundamental à garantia da confidencialidade e integridade dos próprios sistemas informáticos também não é dispensável se a intervenção apenas ocorre quando é "necessária" para permitir o monitoramento e a captura das telecomunicações de forma não criptografada[123]. Em particular, a proteção mais rígida para sistemas de tecnologia da informação não é removida pelo fato de que uma medida de monitoramento de telecomunicações não pode ser realizada com sucesso sem tais intervenções.

A proteção também não é invalidada pelo fato de que a interferência pode ser posteriormente revertida[124]. Se a integridade e a confidencialidade dos sistemas de tecnologia da informação forem ameaçadas ou mesmo afetadas pela interferência, então a proteção dos direitos fundamentais é ativada sem que isso possa ser revertido pela eliminação posterior das consequências da ingerência – para além do fato de que, do ponto de vista técnico, de acordo com as declarações dos peritos ouvidos pelo Tribunal Constitucional, um restabelecimento total do status quo ante não deve ser viável. Portanto, requisitos especiais materiais e procedimentais também são necessários para um monitoramento de telecomunicações na fonte. Normas como o art. 100a do Código de Processo Penal, que foram criadas para a vigilância das telecomunicações tradicionais, também não contêm autorizações para intervenções dessa intensidade particular; eles também não têm uma limitação para um monitoramento na fonte "puro", ou seja, uma salvaguarda legal de que o monitoramento das telecomunicações será limitado à comunicação em curso e que isso será tecnicamente garantido[125]. Com

123 No entanto, esta é a redação do art. 20 l, parágrafo 2, cláusula 1, nº 2 do Projeto de Lei do BKA, BR-Drs. 404/08 AC 5 de junho de 2008.

124 A aceitação por trás da Seção 20 l, § 2º, cláusula 2 c/c a Seção 20k, § 2, cláusula 1, nº 2 do Projeto de Lei BKA parece ser diferente (nota 119).

125 Isto não é referenciado – Bär MMR 2008, 326.

base neles, não são observados os requisitos das limitações que o Tribunal Constitucional Federal formulou sobre o direito fundamental à garantia da confidencialidade e integridade dos próprios sistemas informáticos.

CONCLUSÃO

Em síntese, pode-se afirmar que o Tribunal Constitucional Federal, ao destacar uma necessidade especial de proteção dos sistemas de informática de uso próprio, tem respondido a um potencial de risco particular decorrente do desenvolvimento da informática, de constelações de rede, de muitos novos serviços e das possibilidades de infiltração e manipulação com base neles. O objetivo da proteção continua sendo a proteção da personalidade como base para o desenvolvimento autodeterminado. O tribunal afirmou uma necessidade fundada constitucionalmente de uma salvaguarda especial da confidencialidade e da integridade de sistemas de tecnologia da informação complexos e de uso próprio, que são especialmente importantes para a liberdade de desenvolvimento pessoal nas condições atuais, nas quais a pessoa afetada confia sem esperar poder controlá-los. A proteção oferecida é dirigida contra influências no próprio sistema de tecnologia da informação, mas também abrange a coleta e posterior utilização dos dados por meio de uma infiltração correspondente no sistema de tecnologia da informação. A Constituição não exige uma proibição estrita de tais influências, mas as vincula a requisitos especiais de natureza substantiva e processual.

Em sua decisão sobre buscas online, o Tribunal Constitucional não criou um novo direito fundamental, mas concretizou o direito fundamental há muito reconhecido à proteção da privacidade por meio de uma outra diferenciação. Nesse contexto, o Tribunal, que deve ser cauteloso quanto ao obiter dicta, não pôde se posicionar sobre todas as questões ainda em aberto. A jurisprudência e a ciência do Direito, mas também o legislador, são agora chamados a elaborar os demais contornos da proteção do direito fundamental.

POSFÁCIO

O direito fundamental à garantia da confidencialidade e da integridade dos sistemas de tecnologia da informação formulado pelo Tribunal Constitucional Federal (BVerfG) em 2008, conforme apresentado em meu artigo agora traduzido, já é parte integrante e sedimentado no sistema jurídico alemão. Inicialmente, foram manifestadas críticas a essa construção

elaborada pelo Tribunal[126]. Hoje, existe uma ampla aprovação a este respeito. A jurisprudência dos Tribunais e a literatura científico-acadêmica, bem como a prática e os legisladores têm seguido neste sentido o Tribunal Constitucional alemão. Foram publicadas várias monografias sobre este direito fundamental que tratam exclusivamente sobre ele[127], mas igualmente muitas outras obras que o fazem em relação a outros temas, bem como um grande número de artigos científicos e discussões nos comentários à Lei Fundamental.

Como resultado do desenvolvimento da transformação digital e do uso de sistemas de tecnologia da informação em praticamente todas as áreas da vida, cresceu a consciência de que a interferência nos sistemas de tecnologia da informação pode ter consequências especialmente graves. Seria falta de visão tomar apenas precauções contra intervenções específicas, como medidas concernentes a pesquisas individuais. Essa proteção seletiva deixaria lacunas consideráveis na proteção. A este respeito, faz sentido focar na proteção do sistema, ou seja, em particular na funcionalidade técnica e social dos sistemas de tecnologia da informação, como um pré-requisito para seu uso autônomo para diferentes fins.

Em 2016, o Tribunal Constitucional Federal, além das afirmações de 2008, deixou claro em decisão de revisão da constitucionalidade da lei da Polícia Criminal Federal[128] que os sistemas de informática protegidos não incluem apenas os computadores pessoais dos afetados, mas também aqueles que estão em rede com sistemas de TI de terceiros que funcionam com computadores, por exemplo, na utilização das assim chamadas nuvens[129]. O Tribunal enfatizou expressamente que os dados que são terceirizados para servidores externos com uma confiança legítima na confidencialidade são cobertos pela proteção. Esta é uma reação clara aos perigos associados às possibilidades de aplicação ampliadas e, acima de tudo, à rede de tecnologias digitais.

126 Ver a nota 66 do meu artigo.

127 WEHAGE, Jan-Christoph. O direito fundamental de garantir a confidencialidade e integridade dos sistemas de tecnologia da informação e seus efeitos no direito civil (2013); HEINEMANN, Marcus. Proteção dos sistemas de tecnologia da informação sob direitos fundamentais: com atenção especial ao direito fundamental de garantir a confidencialidade e integridade dos sistemas de tecnologia da informação (2015); HAUSER, Markus. The IT Basic Right: Interfaces and Effects (Duncker & Humblot, 2015); TARAZ, Daniel. O direito fundamental à garantia da confidencialidade e integridade dos sistemas informáticos e à garantia da privacidade digital no âmbito dos direitos fundamentais: preparando o caminho para o digital.

128 BVerfGE 141, 220, 303 e ss.

129 BVerfGE 141, 220, 304.

Uma vez que o novo direito formulado – geralmente denominado direito fundamental computacional – tem a sua base constitucional nos arts. 1º e 2º da Lei Fundamental (proteção da dignidade humana e proteção da personalidade), a garantia dos direitos fundamentais derivados destas normas contém ambas as garantias jurídico-subjetivas e jurídico-objetivas da proteção[130]. O nível jurídico-objetivo da proteção dos direitos fundamentais visa moldar o sistema jurídico de proteção da liberdade e, portanto, é dirigido, em particular, como mandato ao legislador federal e estadual, para tomar medidas de proteção no sistema jurídico. O Tribunal Constitucional Federal enfatiza que a função de proteção jurídico-objetiva não apenas vincula o Estado, mas também afeta a relação entre os particulares. Segundo o entendimento jurídico alemão, isso ocorre no decurso do chamado efeito indireto de terceiros ou efeito horizontal dos direitos fundamentais. Como resultado, a vinculação dos direitos fundamentais também pode entrar em vigor em disputas de direito civil[131].

O Tribunal Constitucional acentua que os efeitos das possibilidades técnicas de processamento de dados estão se tornando cada vez mais importantes para o relacionamento entre particulares. Os serviços básicos para o público em geral com base em extensas coletas de dados pessoais e medidas de processamento de dados seriam fornecidos por empresas privadas, muitas vezes poderosas. Estas tiveram uma influência duradoura na formação da opinião pública, na atribuição e negação de oportunidades, na participação na vida social e nas atividades elementares da vida cotidiana. Tendo em vista a possibilidade de manipulação, reprodução e a possibilidade temporal e espacialmente praticamente ilimitada de disseminação de dados, bem como sua recombinação imprevisível em processos não transparentes, os cidadãos individuais ficaram em dependências de longo alcance. A Constituição alemã oferece proteção contra isso.

O Tribunal vai um passo além e aponta que o impacto do direito fundamental na área de ação privada é particularmente importante se as empresas privadas passarem para uma posição dominante semelhante à do Estado ou se assumirem a provisão do quadro para o setor público de comunicação. Neste contexto, a vinculação dos direitos fundamentais de indivíduos privados seria próxima ou igual a uma vinculação dos direitos funda-

130 BVerfGE 152, 152, número de margem 85 – 88.

131 Aqui e para a sequência BVerfGE 152, 152, número de margem 85.

mentais do Estado em específico[132]. Estas últimas declarações não se referem especificamente à proteção dos sistemas de tecnologia da informação, mas devem ser entendidas de forma que tais empresas também tenham que observar as obrigações de proteção dos direitos fundamentais a esse respeito.

Em outra decisão de 8 de julho de 2021[133], o tribunal especificou a obrigação do Estado de fornecer proteção em mais detalhes, no sentido de que o Estado tem a obrigação de ajudar a garantir que a integridade e confidencialidade dos sistemas de tecnologia da informação sejam protegidos contra ataques de terceiros[134]. Este mandato de proteção é condensado em uma obrigação de proteção concreta sob os direitos fundamentais se o Estado estiver ciente das lacunas de segurança nos sistemas de tecnologia da informação que podem ser usadas por terceiros para se infiltrar nesses sistemas e pesquisar as informações neles encontradas. Ao acessar todo o banco de dados de um sistema de tecnologia da informação, este também pode ser manipulado e os perpetradores podem ameaçar de modo extorsivo a manipulação, em particular mediante a destruição de dados. Voltarei a isso em um momento posterior. Essa aplicação de proteção concentrada é acionada em particular pelo alto risco e potencial de danos das brechas de segurança.

A Corte Constitucional alemã também justifica a necessidade de proteção daqueles que confiam os dados a sistemas de tecnologia da informação com o fato de que os indivíduos são frequentemente dependentes de tais sistemas, e, portanto, a suposição é irreal de que eles poderiam evitar a espionagem, abstendo-se de usar meios digitais de comunicação[135].

Para além deste dever de proteção, o Tribunal destaca ainda que, em casos excepcionais, pode justificar-se que os órgãos do Estado autorizem a existência de lacuna de segurança de que são conhecidos e, por sua vez, o acesso aos sistemas de dados deste sistema. No entanto, isto só é permitido a título de exceção, designadamente com o objetivo de evitar perigos particularmente graves, em especial os perigos do terrorismo internacional. Para fazer uso dessa exceção, é necessária uma autorização legal expressa, na qual são elencados requisitos de justificação adicionais.

132 BVerfGE 152, 152, número de margem 88.

133 BVerfG, decisão de 08.06.2021, Beck RS 2021, 19234, número de margem 26 e ss.

134 Op. cit., número de margem 35.

135 Op. cit., número de margem 33.

Já mencionei que a necessidade de proteção é desencadeada pela expansão dos potenciais de risco associados às tecnologias digitais. A inteligência artificial amplia o potencial de perigos, mas, por outro lado, também contém possibilidades de reconhecer e combater tais ataques. Fala-se, neste contexto, do potencial de uso duplo da IA. O âmbito da proteção do direito fundamental à integridade e à confidencialidade dos sistemas de tecnologia da informação também inclui ameaças que recentemente chamaram particular atenção. Refiro-me ao uso direcionado de spyware, como o software Pegasus. Isso permite, entre outras coisas – como ficou conhecido em 2021[136] –, o monitoramento remoto de smartphones. Para tanto, são utilizadas lacunas de segurança no software, algumas das quais foram criadas por solicitação expressa dos usuários do software ou foram deliberadamente deixadas após serem detectadas. O spyware suportado por IA tem sido usado em grande escala por várias instituições (privadas, mas também estatais estrangeiras) para espionagem ilegal, em particular de políticos de alto escalão, ativistas de direitos humanos e jornalistas. O fabricante israelense da tecnologia de vigilância – o Grupo NSO – confiou no fato de que a empresa proíbe os compradores do software de uso ilegal e sanciona-o caso se torne conhecido. Isso não parece ter sido muito eficaz.

O problema é particularmente sério quando o "sequestro" de sistemas de tecnologia da informação pode ter consequências para grandes partes da sociedade, por exemplo, paralisando o fornecimento de energia ou água, ou mesmo interrompendo cadeias importantes de abastecimento[137]. Os sistemas de tecnologia da informação usados na produção industrial também podem ser afetados. Mais recentemente, foram conhecidos casos em que sistemas de tecnologia da informação foram hackeados para exigir um alto valor de resgate pela "descriptografia" do software malicioso, que também foi pago por um alto valor.

Os militares também dependem de sistemas de tecnologia da informação. A infiltração em seus sistemas de tecnologia da informação pode, por exemplo, bloquear infraestruturas militarmente importantes ou prejudicar a funcionalidade dos sistemas de armas.

136 Cf., por exemplo, as reportagnes no Jornal Süddeutschen de 20.07.2021, nº 164, S. 9 – 11 assim como do dia 21.07.2021, nº 165, p. 1, 8 e ss.

137 Um exemplo é o ataque perpetrado no ano de 2021 Supply-Chain, que foi efetivado no software do servidor Casey, provavelmente pelo grupo Hacker REvil.

São acréscimos que não estão diretamente relacionados às questões tratadas no ensaio monográfico. Mas, mesmo no momento do seu desenvolvimento, é necessário demonstrar – como em 2008 – que os direitos fundamentais devem ser interpretados de forma dinâmica. Isso significa que eles podem ou mesmo devem reagir às mudanças em sua área real – aqui, às mudanças nas tecnologias e às novas ameaças associadas. A proteção do sistema se tornou cada vez mais importante em tempos de transformação digital. O Estado deve proteger esses sistemas, determinar que os particulares o façam, bem como monitorar o cumprimento dessas obrigações pela autoridade pública.

Sobre o autor:

Wolfgang Hoffmann-Riem

Professor afiliado de Inovação e Direito na Bucerius Law School, Hamburgo. Professor Emérito de Direito Público e Administração Pública da Universidade de Hamburgo. De 1995 a 1997, foi Chefe do Departamento de Justiça (Senador) do Estado de Hamburgo, além de Presidente do Comitê de Direito do Bundesrat alemão. De 1999 a 2008, foi Juiz do Tribunal Constitucional alemão. Seu campo de responsabilidade como Relator incluiu – entre outros – a proteção da privacidade e dos dados, a liberdade de expressão e informação, bem como a inviolabilidade do lar. De 1979 a 1995, 1997 a 1999, foi Diretor do Hans-Bredow-Insitute on Radio and TV-Broadcasting. Na Universidade de Hamburgo, fundou e presidiu o Centro de Pesquisa em Direito e Inovação (1995-2012). Ele ainda é um dos Diretores do Instituto de Pesquisa sobre a Lei de Proteção Ambiental da Universidade de Hamburgo. Desde 2007, é Membro alemão da "Comissão Europeia para a Democracia através do Direito" (Comissão de Veneza) do Conselho da Europa.

Sobre o tradutor e a revisora técnica:

Italo Roberto Fuhrmann

Doutorando em Direito pela PUCRS. Advogado.

Jacqueline de Souza Abreu

Doutoranda em Direito na Faculdade de Direito da Universidade de São Paulo. Advogada. Mestra em Direito pela University of California, Berkeley (EUA), com foco em Direito e Tecnologia, e pela Ludwig-Maximilians-Universität München (Alemanha), com foco em Direitos Fundamentais. Graduada em direito pela Universidade de São Paulo. Cocoordenadora do Dossiê "Privacidade e Proteção de Dados Pessoais na Segurança Pública e no Processo Penal", da *Revista de Direito Público*, do IDP.

Artigo convidado.

Privacy, Data Protection and Law Enforcement. Opacity of the Individual and Transparency of Power[1]

SERGE GUTWIRTH[2]
Vrije Universiteit Brussel (Bélgica).

PAUL DE HERT[3]
Vrije Universiteit Brussel (Bélgica).

SUMMARY: Introduction; 1 Principles of the democratic constitutional state; 1.1 The Recognition of Human Rights in their Double Function; 1.2 The Rule of Law; 1.3 Democracy; 2 The democratic constitutional state and the invention of two complementary legal tools of power control; 2.1 Limiting power through opacity tools; 2.2 Channelling power through transparency tools; 3 Privacy as a tool for opacity (creating zones of non-interference); 3.1 The negative role of privacy; 3.2 The positive role of privacy; 3.3 The non-absolute nature of privacy; 4 Data protection as a tool for transparency; 4.1 Introduction; 4.2 The rationale behind data protection; 4.3 Data protection as an opacity tool?; 4.4 The charter of fundamental rights of the european union; 5 The shift from opacity towards transparency in european human rights law; 5.1 European human rights law and the legality requirement; 5.2 The success of the legality requirement; 5.3 A critical comment about the strasbourg focus on the legality requirement; 5.4 The danger of proceduralisation; 5.5 A requirement fundamental to opacity: necessary in a democratic state; 6. Combining privacy and data protection 6.1 Combining the tools; 6.2 Determining the switch; 6.3 An example: camera surveillance; 6.4 A second example: passenger profiling; 6.5 Workable criteria?; Conclusion.

1 Original Version available in: DE HERT P. & S. GUTWIRTH, "Privacy, data protection and law enforcement. Opacity of the individual and transparency of power" in E. CLAES, A. DUFF & S. GUTWIRTH (eds.), *Privacy and the criminal law*, Antwerp/Oxford, Intersentia, 2006, pp. 61-104. (ISBN 90 5095 545 2).

2 For Serge Gutwirth this contribution is partly the result of research carried out under the *Interuniversity Attraction pole* V.16 research *The loyalties of knowledge* financed by the *Belgian Federal Science Policy* (see: www.imbroglio.be).

3 For Serge Gutwirth this contribution is partly the result of research carried out under the *Interuniversity Attraction pole* V.16 research *The loyalties of knowledge* financed by the *Belgian Federal Science Policy* (see: www.imbroglio.be).

INTRODUCTION

Privacy constitutes a relatively new concept in the development of contemporary law. Its beginnings are traditionally attributed to the famous publication of an article "The right to privacy" by the American scholars Warren and Brandeis in the *Harvard Law Review* at the end of the 19th century[4].The piece was a reaction against the state of American journalism, wherein they complained about the journalists' lack of respect for personal feelings and sexual relations. The authors called for privacy which they defined as the right to be let alone. Although notions such as secrecy and confidentiality were undeniably at the heart of the concept of privacy, there is a strong tendency in case law and literature to understand privacy as a broadly conceived concept of autonomy and information autonomy of the human person[5]. In this view, privacy embodies the freedom of choice, autonomy and self-determination of individuals in social and relational matters.

In Europe, the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR), concluded in 1950, plays a crucial role regarding the protection of the right to privacy contained in Article 8 of the Convention. During the last few decades the European Court of Human Rights in Strasbourg, which has the power to make rulings about violations of art. 8 ECHR, has developed a vast and relevant but not unambiguous body of case-law about privacy.

Today, however, privacy has come increasingly under pressure, not only as a result of the large scale introduction of information technology and monitoring techniques, but also as a result of the far reaching public security policies that were devised in the aftermath of September 11. These evolutions have far reaching consequences for law enforcement. Existing principles of criminal law are challenged and traditional safeguards for defendants are threatened. In its report *Security and Privacy for the Citizen in the Post-September 11 Digital Age*, the Institute or Prospective Technological Studies (EU-Joint Research Centre) came to the conclusion that "(t)he move towards proactive surveillance reflects a transformation from the traditional legal model of gathering conclusive evidence of wrongdoing 'beyond reasonable

4 S. D. WARREN & L. D. BRANDEIS, "The right to privacy", *Harvard Law Review*, 1890, pp. 193-220.

5 E. G. S. GUTWIRTH, *Privacy and the information age*, Lanham, Rowman & Littlefield Publishers, 2002, 158 p. and Rigaux, F., *La protection de la vie privée et des autres biens de la personnalité*, Brussels/Paris, Bruylant/L.G.D.J, 1990, 849 p.

doubt' to put before a criminal court, towards an intelligence-gathering and disruptive model". This partly reflects the concerns of intelligence agencies and proactive squads to protect their information sources from disclosure to defendants in the criminal justice process, but partly a methodological shift to crime risk reduction, using probabilistic profiles to stop criminals (especially terrorists) from reaching their targets, concluding business deals or purchasing businesses/property to integrate into the upper world. This then raises important questions for the preparedness of the criminal justice systems of Member States and court-oriented protections for citizens. "What will the role of the courts be, and how will the use of secret intelligence to guide police actions be treated for disclosure to defendants to protect their rights?"[6].

These findings and questions are essential for our present study of privacy and criminal procedure. The evoked shifts in policing strategies have been opposed by civil rights movements and legal authorities precisely with privacy arguments. The tone of the debate has often been rather dark, with many doom scenario's and catchphrases such as *The Death of Privacy* or *The End of Privacy...*

We believe that there is no reason why these scenarios should become an European reality. Europe has a long, stable tradition of dealing with power demands and threats to human rights. To discover this tradition we first look into the constitutional role that privacy plays in our liberal democratic constitutional state, thereby distinguishing between privacy and data protection. For us privacy is an example of a "tool of opacity" (stopping power, setting normative limits to power), while data protection and criminal procedure can be mainly – not exclusively – seen as "tools of transparency" (regulating and channelling necessary/reasonable/legitimate power). Much can thus be learned from making and ascertaining the differences in scope, rationale and logic between privacy on the one hand, and data protection on the other. This is, in two short sentences, both the program and the challenge of this contribution.

New threats to privacy should and can be addressed by using measures relating to privacy *and* data protection. Examples such as video surveillance

6 Institute For Prospective Technological Studies – Joint Research Centre, "*Security and Privacy for the Citizen in the Post-September 11 Digital Age. A prospective overview*. *Report to the European Parliament Committee on Citizens" Freedoms and Rights, Justice and Homme Affairs* (LIBE), July 2003, IPTSTechnical Report Series, EUR 20823 EN, 97 (our emphasis). See also: ftp://ftp.jrc.es/pub/EURdoc/eur20823en.pdf.

and data gathering by the U.S. Department of Homeland Security (CAPPS-II) show how privacy and data protection have their respective and different roles to play. We critically assess the case-law of the European Court of Human Rights. In its current formulation, the privacy doctrine by the European Court neglects the importance of opacity in a democratic state. The European judges prefer to focus on much safer issues such as accountability and foreseeability, whereas our times are in need of more or less clear-cut statements about the reasonableness of new police powers that are developed within and outside Europe.

1 PRINCIPLES OF THE DEMOCRATIC CONSTITUTIONAL STATE

The project of the democratic constitutional state has engendered alternative ways of dealing with power and brought about a specific concept of the state in which power is by definition limited. The aims of the democratic constitutional state are generally translated, expressed and concretised through the enactment of three basic constitutional principles, namely the recognition of fundamental rights and liberties, the rule of law (constitutionalism) and democracy. These three principles will be briefly discussed hereafter[7].

1.1 The recognition of human rights in their double function

First, the constitutions of democratic constitutional states recognise a set of individual fundamental rights and freedoms (or shortly: human rights) that are deemed to be at the very core of the political construct[8].

7 This chapter includes parts of P. DE HERT & S. GUTWIRTH, "Making sense of privacy and data protection. A prospective overview in the light of the future of identity, location based services and the virtual residence" in Institute For Prospective Technological Studies – Joint Research Centre, *o.c.*, pp. 111-162. Available at: ftp://ftp.jrc.es/pub/EURdoc/eur20823en.pdf. See also S. GUTWIRTH, "De polyfonie van de democratische rechtsstaat" [The polyphony of the democratic constitutional state] in *Wantrouwen en onbehagen*, M. ELCHARDUS (editor), Balans 14, Brussels, VUBPress, 1998, pp. 137-193.

8 According to the original political philosophy of the Enlightenment (Locke, Rousseau, [...]) these rights are *natural rights*. This view sees individuals as the holders of inalienable rights because of the mere fact that they are humans. This implies that the individual exists as the bearer of a number of eternal and universal rights which transcend and bind the (temporal) power of the state. One step further, it is argued that if a government, does not respect these rights, it becomes illegitimate and the people have the right to resist and to abolish it (cf. LOCKE, the *Declaration of Independence* and the *Déclaration des droits de l'homme*). We are not willing to enter the debate about the theory of natural rights. Elsewhere, however, and inspired by Rawls, we defend a non-comprehensive and non-essentialist, but political and constructivist (pragmatic) approach on human rights and basic liberties; cf. P. DE HERT & S. GUTWIRTH, "Rawls" political conception of rights and liberties. An unliberal but pragmatic approach to the problems of harmonisation and globalisation" in M. VAN HOECKE (Editor) *Epistemology and Methodology of Comparative Law*, Hart Publications, Oxford/Portland, 2004, pp. 317-357.

In principle, the State is not allowed to encroach upon or to interfere with these rights. Human rights work as a shield or a bulwark. They express the recognition of the power of the individual, drawing the limits and frontiers of the power of the state and of state intervention[9]. Hence, individuals have acquired a package of elementary prerogatives against the state power[10]. For the understanding of privacy and its implications, it is crucial to bear in mind that this recognition of human rights affirms the existence of the human individual as an independent being, detached from the state but also from the politics driven by a democratically elected majority. Human rights protect individuals against J. S. Mill's proverbial "tyranny of the majority". In other words, they create a sphere of individual autonomy or self-determination and in doing so, they protect individuals against excessive steering of their lives; they contribute to the creation of the private sphere[11].

Human rights and liberties not only restrict the power of the state, but also empower citizens (or individuals) to participate in the political system[12]. This second important function of human rights explains why within the Western political tradition, it may not be too hard to find an overlapping consensus on the importance of basic liberties, such as freedom of expression, liberty of conscience and freedom of association. These rights and liberties enable citizens to develop and exercise their moral powers in forming, revising and in rationally pursuing their conceptions of the good[13].

---

9 This is easy to understand as the revolutionary people of the 17/18th century in England (1688 Bill of Rights), the USA (1776 Declaration of Independence and 1788-1791 *US Constitution*) and France (1789 *Déclaration*), which laid the building stones of the democratic constitutional state, were simultaneously doing two things. On the one hand, they did put an end to the arbitrary and absolute power of the former sovereigns and, as such, the described (r)evolutions can be said to have broken with the political economy of the past. On the other hand the actors of the three (r)evolutions laid the basis for a new constitutional system in order not only to make absolutism and arbitrary power relations impossible in the future but also to build up a new kind of society wherein individual liberty and entrepreneurship would prevail.

10 Actually this is also, but indirectly, the case pertaining to the power of other societal actors.

11 Of course, it must be said that this sphere of autonomy and self-determination is not an absolute one: it is subject to limitation, through legal mediation processes of legislative and/or judicial nature, in name of the protection of the rights and interests of the other citizens or of the public interest. This is indeed the case of privacy, which, beyond the fact that it is an quintessential value in a democratic constitutional state, is nonetheless not offering an absolute protection to its holder (cf. *infra*).

12 C. SCHNEIDER, "The Constitutional Protection of Rights in Dworkin's and Habermas 'Theories of Democracy'", *UCL Jurisprudence Review*, 2000, 118 with reference to Charles Larmore ("The Foundations of Modern Democracy: Some Remarks on Dworkin and Habermas", *European Journal of Philosophy*, 1995, 65) who has observed that "individual rights serve, not to protect us against the collective will, but rather to protect the means necessary for creating a collective will". Cf. P. DE HERT & S. GUTWIRTH, "'Rawls' Political Conception of Rights and Liberties", *l.c.*, pp. 317-357.

13 J. RAWLS, *Justice as Fairness. A Restatement*, Cambridge, Harvard UP, 2001, § 13.4.

1.2 The rule of law

Secondly, the constitutions of democratic constitutional states all enshrine the rule of law and constitute a *Rechtsstaat*. The constitutional recognition and implementation of the rule of law again tend to limit the power of government, but this time this happens no longer through setting a limit to the reach of the power, but through what one could call a system of "internal" organisation of government and power. Nonetheless, the objective remains the same, namely the protection of individuals against excessive and arbitrary domination. The main idea of the rule of law is the subjection of government and other state powers to a set of restricting constitutional rules and mechanisms[14].

On the one hand the rule of law provides for the principle of legality of government, which stands for the basic principle that power can only be exercised in accordance to the law. From this perspective public authorities are bound by their own rules and can only exercise their powers in a lawful way. All powers must derive from the constitution (which in its turn is deemed to translate the will of the sovereign people) and any exercise of power must be interpretable as emanating from a constitutional provision. This implies the important fact that the government is accountable and that its actions must be controllable, and thus transparent. "The rule of law" thus refers to the idea that our societies are governed by rational and impersonal laws and not by the arbitrary commands of humans. Moreover, because these laws must be general and apply to all, they (at least formally) embody the principle of equal treatment and protection of the laws[15].

On the other hand the rule of law establishes the trias politica or, in other words, a system of balancing of powers. Here the basic idea is to limit the power of the state by spreading it over different centres, with different competencies and functions. These powers – the executive, legislative and judicial power – are constitutionally doomed to work together through a dynamic system of mutual control or checks and balances. This system relies heavily on the famous ideas which Montesquieu developed in *De l'esprit des lois*: "Pour qu' on ne puisse abuser du pouvoir, il faut que, par la disposition des choses, le pouvoir arrête le pouvoir"[16]. Indeed, the best way to limit

14 J. CHEVALLIER, *L'Etat de droit*, Paris, Montchrestien, 1994, p. 11.
15 K. RAES, *Tegen betere wetten in. Een ethische kijk op het recht*, Gent Academia Press, 1997, p. 215.
16 MONTESQUIEU, *De l'esprit des lois*, t. 1, Paris, Garnier-Flammarion, 1979, p. 293.

power is to divide it up and to spread it over competing centres. In sum: the trias politica replaces a centralist power by a pluricentric power economy[17]. Such a system implies the mutual accountability of state powers, and thus again, the reciprocal transparency and controllability of the legislative, the judicial and, last but not least, the executive power.

1.3 Democracy

Thirdly, the constitutions of democratic constitutional states recognise the postulate of the people's sovereignty and the principles of democracy and democratic representation. During the political Enlightenment the sovereignty of the rulers gave way to the people's sovereignty and the idea of the political self-determination of the nation. Consequently, in a democratic constitutional state the only valid justification of power must be sought in the citizens' consent or will. This crucial link is expressed through the different variations upon the theme of the social contract (Beccaria, Locke, Rousseau [...]), for such contracts construe the constitution of a political entity with reference to the will or consent of the individuals. State powers are derived from the sovereignty of the citizens.

More concretely, these theoretical foundations of state power imply that government in all its aspects must be in line with the public or general interest and must mainly be driven by the will of the majority. Hence, systems of representation and participation of citizens are of crucial importance. State organs and institutions must be representative. Participation by citizens in political decision-making must be organised and stimulated. And, last but not least, systems of democratic governance must provide procedures for the direct and indirect control of the public authorities by the citizens. As a result democratic rule implies the accountability of the government towards the citizens, which again calls for transparency of public decision-making and policies.

The foregoing analysis can be summarised by highlighting the fact that the development of the democratic constitutional state has led to the invention and elaboration of two complementary sorts of legal tools which

17 R. FOQUE, "Rechtsstatelijke evenwichten in de trias politica. De actuele betekenis van de onafhankelijkheid van de rechterlijke macht", *Vigiles – Tijdschrift voor politierecht*, 1996/4, pp. 1-5 and R. FOQUE, "Rechtsstatelijke vernieuwing. Een rechtsfilosofisch essay" in P. KUYPERS, R. FOQUE & P. FRISSEN, *De lege plek van de macht. Over bestuurlijke vernieuwing en de veranderende rol van de politiek*, Amsterdam, De balie, 1993, pp. 18-44.

both aim at the same end, namely the control and limitation of power. We make a distinction between on the one hand tools that tend to guarantee non-interference in individual matters, or the opacity of the individual, and on the other, tools that tend to guarantee the transparency/accountability of the powerful[18]. This will enable us, in the following sections, to link privacy to the first tool and data protection to the second.

2 THE DEMOCRATIC CONSTITUTIONAL STATE AND THE INVENTION OF TWO COMPLEMENTARY LEGAL TOOLS OF POWER CONTROL

2.1 Limiting power through opacity tools

As the core aim of a democratic constitutional state is to foster an order driven by individual liberty, the protection of individuals against state interference is at premium. Hence, the importance of the private sphere as a sphere where individual liberty has a privileged status cannot be underestimated. That is why opacity tools that shield individuals against state interference are so crucial.

The ideas behind such tools can be understood by recalling the function of the first generation of human rights. By recognising human rights, the revolutions of the 17th-18th centuries in England, the US and France laid the foundations for a sharper legal separation between the public and private spheres[19]. The constitutional recognition of these rights led to the creation of a sphere of individual autonomy and self-determination, where the citizens may live their lives without interference of the state. Human rights have empowered the individuals through recognition of their liberty and prerogatives. And inversely, limits to state power were drawn through the recognition of the autonomy of the citizens.

Thus, human rights can be understood as legal tools that protect individuals against interference by the state and by private actors. They tend to require abstention from undesired intervention in matters that are essential for the protection of the individual's autonomy and liberty.

18 "Opacity" designates a zone of non-interference which in our opinion must not be confused with a zone of invisibility: privacy for instance does not imply secrecy, it implies the possibility of being oneself openly without interference. Another word might have been "impermeability" which is too strong and does not contrast so nicely with "transparency" as "opacity" does.

19 See Ph. ARIES & G. DUBY G. (editors), *Histoire de la vie privée* and particularly M. Perrot (editor), *Volume 4: De la Révolution à la Grande Guerre,* Paris, Le Seuil, 1987, p. 637.

A good example is the protection of the "sanctity" or inviolability of the home, which indeed properly expresses the concern for respect for individual autonomy: public authorities (but also other citizens) must respect the bounds of the home. A home is inviolable, and any breach of that principle generally engenders criminal prosecution. Once inside their home, people are more free from interference from the government (and others) than they are outside. A home is a privileged setting. Within a home, each and every person has the freedom to do as he/she pleases, uninhibited by society's social and ethical mores. U.S. case law has already shown, *e.g.*, that watching pornography at home and possessing obscene movies, which may not be distributed in public, is protected by the inviolability of the home. Providing home entertainment by serving food naked cannot be outlawed in the same way than such entertainment in bars and restaurants is. This does not mean that everything happening inside the home is automatically protected. Search warrants can be ordered in criminal cases, but only, in principle, if a series of stringent conditions are met[20]. Crimes and unlawful acts are not condoned because they happen to take place within a home. But because a home is granted a special measure of protection, trespassing by third parties and especially by the police and judicial authorities is strictly regulated.

The relationship between opacity tools and individual liberty is far from simple. The latter is undeniably the prime concern of the former, but the relationship is not without ambiguity, for opacity tools supersede individual consent when societal interests are at stake[21]. What is essential to opacity

20 The political function of the right to have the house protected and the idea that limitations are still possible, have been recognised by the European Court in the *Niemietz* judgement: "More generally, to interpret the words 'private life' and 'home' as including certain professional or business activities or premises would be consonant with the essential object and purpose of Article 8 (art. 8), namely to protect the individual against arbitrary interference by the public authorities (see for example: *Marckx v. Belgium*, judgement of 13 June 1979, § 31). Such an interpretation would not unduly hamper the Contracting States, for they would retain their entitlement to 'interfere' to the extent permitted by paragraph 2 of Article 8 (art. 8-2); that entitlement might well be more far-reaching where professional or business activities or premises were involved than would otherwise be the case"; *Niemietz v. Germany*, judgement of 16 December 1992, § 32). *Note about our references to the judgements of the European Court of Human Rights*: in a first reference to such a judgement we will mention the name of the applicant and the respondent state, as well as the date of the judgement. With these data the judgements can easily be found at: http://www.dhcour.coe.fr/eng/. After the first reference that we will only use the name of the applicant (ex. *Niemietz*). For precise references and quotes we will refer to the relevant paragraphs of the judgements.

21 A fine example is found in the provisions in the Directive no. 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data on the free movement of such data, *Official Journal of the European Communities*, L 281, 23 November 1995, 31-50 (further cited as Data Protection Directive) establishing a public law regime that cannot be varied by a private law contract. Consequently, any agreement pursuant to which a data subject waives some or all of his rights under the Data Protection Directive is void and unenforceable, even if the agreement otherwise meets

tools is their normative nature. Through these tools, the (constitutional) legislator takes the place of the individual as the prime arbiter of desirable or undesirable acts that infringe on liberty. This collective, normative dimension of opacity tools explains the complex relationship between human rights and individual liberty. The harm principle as a yard stick to measure wrongful infringements on individual liberty is replaced by a more formal criterion[22], and ad hoc balancing is replaced by categorical balancing[23]. The stated goal of human rights, even in cases when they limit freedom by taking away consent, is the protection of individual liberty. Through the implementation of these tools, the outcome of conflicts becomes more predictable: the balancing of (conflicting) interests has already been done in favour and in the name of the subject. One author holds that "a right defining liberty often offers greater protection than the liberty itself"[24].

It is however evident that categorical balancing does not always result in the greatest possible protection of the individual, for instance when (constitutional) texts provide for intrusions into the home when a search warrant has been obtained. Therefore, it is more precise to emphasise the normative rather than the protective function of opacity tools. Through such tools the (constitutional) legislator enacts hard or clear norms. Choices about the way liberty interests and other interests should be balanced are made in an abstract way. Complexity is reduced. Of course, complexity can be reduced to the full detriment of liberty, but since this would go against the spirit of the Western constitutional state as defined above, this may not be taken seriously as a hypothesis.

To end this section, we wish to underline that opacity tools do not necessarily take the form of a human rights provision enshrined in an international treaty or national constitution. For instance the existence and intervention of the investigating judge in the criminal process is a good example of an opacity tool, because it expresses a clear or hard choice to grant special status to the protection of the private sphere. However, in common law legal systems there is no such institution. And even the

all validity requirements and is in the data subject's interest. Although data protection, when applied, will be identified as a typical transparency tool (*infra*), it is constructed on public law foundations. The sheer existence of European data protection law in its present (public law) form is a "hard" choice.

22 J. Ravanas, annotation to Cour de Cassation (Fr.), 5 March 1997, *Recueil Dalloz*, (Cahier Jurisprudence), 1998, v. 34, (474-476), 475 with reference to the work of François Rigaux: "La sanction est fondée sur la violation du droit de demandeur, quel que soit le comportement du défendeur".

23 *Ibid.*

24 *Ibid.*: "Que la liberté devienne un droit, la protection en sort renforcée".

constitutions of Continental legal systems (Belgium, the Netherlands, [...]) are silent about the investigative judge, although his position in the criminal process is central. There is properly speaking no human right to have a criminal investigation done by an investigative judge, but there is a legal set of rules that recognises such a positive right[25].

2.2 Channelling power through transparency tools

The second set of tools is connected to the principles of the democratic constitutional state that limit the state powers, not by drawing the limits of their reach through the recognition of a private sphere of autonomy, but by devising legal means of control of these powers by the people, by controlling bodies or organisations and by the other state powers. These tools have the common feature that they are intended to compel government and private actors to "good practices" by focusing on the transparency of governmental or private decision-making and action, which is indeed the primary condition for an accountable and responsible form of governance. The system of checks and balances, for example, installs the mutual transparency of state powers, while the controllability and accountability of government by the citizens implies free and easy access to readily available government information, the enactment of swift control and participation procedures, the creation of specialised and independent bodies to control and check the doings of government, and so on.

The tools of opacity are quite different in nature from the tools of transparency. Opacity tools embody normative choices about the limits of power; transparency tools come into play after these normative choices have been made in order still to channel the normatively accepted exercise of power. While the latter are thus directed towards the control and channelling of legitimate uses of power, the former are protecting the citizens against illegitimate and excessive uses of power. The latter do take into account that the temptations of abuse of power are huge, and empower the citizens and special watchdogs to have an eye even on the legitimate use of power: they put counter powers into place. The former determine what is in principle out of bounds for governmental and private actors and, hence, what is deemed so essentially individual that it must be shielded against public and private

25 P. DE HERT, "Het recht op een onderzoeksrechter in Belgisch en Europees perspectief. Grondrechtelijke armoede met een inquisitoriale achtergrond" [The investigating judge in Belgian and European Law], *Panopticon. Tijdschrift voor strafrecht, criminologie en forensisch welzijnswerk*, 2003, pp. 155-198.

interference. On the one hand there is a regulated acceptance; on the other there is a prohibition rule, which is generally subject to exceptions[26].

This second set of tools is particularly useful for regulating relationships between private actors. As a starting point for such relationships it should be accepted that these actors have equal claims to liberty and are in principle capable of protecting their own liberty interests. Individual consent and ad hoc balancing are suitable instruments to reconcile the liberty interests at stake. Only after careful consideration and with solid arguments, for instance with regard to unequal power relationships, should governments interfere and impose "hard norms" or "choices" resulting from categorical balancing.

3 PRIVACY AS A TOOL FOR OPACITY (CREATING ZONES OF NON-INTERFERENCE)

3.1 The negative role of privacy

Privacy pre-eminently imposes itself as the legal concept translating the political endeavour to ensure non-interference (or opacity) in individual matters. It is embedded in the contemporary democratic constitutional state, the values of individualism and the constitutional separation between state and church. It is also intimately linked with the idea that individuals are able and willing to unshackle themselves from tradition, social conventions or religion and dissociate themselves, up to a point, from their roots and upbringing. Privacy, negatively stated, protects individuals against interference in their autonomy by governments and by private actors[27]. It is a fundamental notion for a project of society that wants to limit power relationships.

The work of the French liberal Benjamin Constant (1767-1830) is illustrative and very important in this respect[28]. Constant can claim

26 In essence the second category of tools overlaps with administrative law promoting and ensuring accountability by governmental actors. Its logic can also be found in Western labour law. Without touching the power relations between workers and employers, this set of rules provides for a kind of procedural justice, viz. rules to be followed when taking certain decisions that cannot be challenged due to differences in power position, but that can be checked for fairness in the decision making.

27 Such a negative understanding of privacy can clearly be read in the formulation of Article 8 ECHR: no interference by public authorities is permitted unless necessary in a democratic society.

28 In particular *Principes de politique*, written between 1806 and 1810 but first fully published in1980: B. CONSTANT, *Principes de politique applicable à tous les gouvernements*, (1806-1810), E. HOFMANN (editor), Hachette, Paris, 1997, p. 447. See on Constant: P. DE HERT, "The Case of Anonymity in Western Political Philosophy. Benjamin Constant's Refutation of Republican and Utilitarian Arguments against Anonymity", in C. NICOLL, J. E. J. PRINS & M. J. M. VAN DELLEN (editors), *Digital Anonymity and the Law.*

fatherhood, not only of the paradigm of the state as a potential menace to individual liberty, but also of terms such as "liberalism", "private life" ("vie privée"), "surveillance", "individual liberty" and "the individual" so dear to modern privacy thinking. Constant saw the recognition of privacy and individual liberty as an unavoidable fact and coupled this from the start to the idea of limited government. This liberal militancy was born out of the spirit of doubt. He was convinced that no group of mortals could be certain about the nature of virtue or the human good and he was ready to oppose any regime that claimed such certainty[29]. No social telos could be so unquestionable as to justify legal enforcement. Governmental attempts to outlaw dissent, legislate public morality, and inculcate civic virtue are paternalistic violations of the purity of man's moral judgement and his capacity to seek for the truth himself[30]. Individuals lose their critical attitude once the truth is imposed upon them. States should limit themselves to their primary task; they should abolish all moral institutions and preserve only the necessary political institutions. Individual liberty should be surrounded with institutional arrangements that keep open and accessible a wide variety of social and political possibilities. Modern times, Constant holds, create a desire for privacy in us (we "must" turn towards the freedom "of peaceful enjoyment and private independence"). Everything that supposes the unnecessary submission of the individual to society should be redefined in the name of individual liberty by applying the harm principle: criminal laws, sexual and religious codes, mores, taxes, state regulation of commerce and industry, state propaganda and access to the media, public schools, army recruitment, etc.

3.2 The positive role of privacy

Privacy also functions positively. Being the legal concept that embodies individual freedom, it plays a quintessential role in the democratic constitutional state based upon the idea that its legitimacy can only result from a maximal respect of each person's individual liberty. Privacy protects the fundamental political value of a democratic constitutional state as it

Tensions and Dimensio, Volume 2 Information Technology & Law Series (IT&Law Series), The Hague, TMC Asser Press, 2003, pp. 47-97.

29 S. HOLMES, *Benjamin Constant and the Making of Modern Liberalism*, New Haven, Yale University Press, 1984, 7.

30 B. CONSTANT, *o.c.*, Book XIV, Chapter III, 313-314; Chapter V, 317. On Constant's defense of self-interest against paternalism, see S. HOLMES, *o.c.*, pp. 252-254.

guarantees individuals their freedom of self-determination, their right to be different and their autonomy to engage in relationships, their freedom of choice, their autonomy as regards – for example – their sexuality, health, personality building, social appearance and behaviour, and so on. It guarantees each person's uniqueness, including alternative behaviour and the resistance to power at a time when it clashes with other interests or with the public interest[31].

In literature the close bond between the negative and positive functions of the right to privacy and its necessity for political life have been rightly stressed. Within Arendt's[32] and Habermas's[33] construction of the public sphere, a space for individuals is provided to develop their own identity and ideas in order to engage in public life. The ideal of a "public" government necessarily entails its opposite: a "private" sphere, protected from public intervention, within which people are free to form individualised relationships that cannot be justified under the requirements of impersonal beneficence[34].

This view is echoed in the available privacy literature[35] and in many anthropological studies. These studies have made it clear that observation

31 About this concept of privacy see a.o. S. GUTWIRTH, *Privacy and the information age*, *o.c.*, *passim* and S. GUTWIRTH "Privacyvrijheid" een voorwaarde voor maatschappelijke diversiteit" [Privacy's freedom: a condition for social diversity] in *Eenheid en verscheidenheid in recht en rechtswetenschap*, A. M. P. GAAKEER & M. A. LOTH (Red), *SI-EUR Reeks 28*, Arnhem, Kluwer/Gouda Quint, 2002, pp. 95-138.

32 Arendt speaks of "the danger to human existence from the elimination of the private realm"; H. Arendt, *The Human Condition.* Chicago: University of Chicago Press, (1958), 1998, 70. She further notes that there are "a great many things which cannot withstand the implacable bright light of the constant presence of others on the public scene", for example, love, and that "a life spent entirely in public, in the presence of others, becomes [...] shallow" (*ibid.*, 71). Arendt therefore recognises that without a private space for identity formation and the shielding of intimate details from others, the public could never be constituted. Indeed she concludes that public and private can "exist only in the form of coexistence" (*ibid.*, p. 59).

33 Compared to Arendt, public sphere theorist Jürgen Habermas carves out a similar yet more powerful role for the private sphere. For Habermas, more than simply coexistent, the private sphere literally constitutes the public. "The bourgeois public sphere may be conceived above all as the sphere of private people who have come together as a public"; J. HABERMAS, *The Structural transformation of the public sphere.* Translated by Thomas Berger and Frederick Lawrence. Cambridge, Mass., MIT Press., 1989, p. 26.

34 L.M. SEIDMAN, "Public Principle and Private Choice", *Yale Law Journal,* Volume 96, 1987, p. 1026.

35 On the positions taken by privacy authors such as Diffie and Landau, Reiman and Bloustein, see Ch. HUNTER, "Political privacy and online politics: how e-campaigning threatens voters privacy", *First Monday,* v. 7, no. 2, (February 2002), 4. Interesting in the light of the Kantian position is the point made by Bloustein: "The man who is compelled to live every minute of his life among others and whose every need, thought, desire, fancy or gratification is subject to public scrutiny, has been deprived of his individuality and human dignity. Such an individual merges with the mass. His opinions, being public, tend always to be conventionally accepted ones; his feelings, being openly exhibited, tend to lose their quality of unique personal warmth and to become the feelings of every man. Such a being, although sentient, is fungible; he is not an individual"; E. J. BLOUSTEIN, "Privacy as an Aspect of Human Dignity", in F. D. SCHOEMAN (editor), *Philosophical Dimensions of Privacy: An Anthology,* New York, Cambridge University Press, 1984, 188. Both Kant and Bloustein ground their argument of the notion of human dignity, with very different outcomes.

by listening or watching which is known to the subject necessarily exercises a restrictive and/or steering influence over him: he must either bring his actions within the accepted social norms in the particular situation involved or decide to violate those norms and accept the risk of reprisal[36].

The significant role of privacy, instrumental to the building of the citizen, should also be understood in the light of Michel Foucault's that all power relationships presuppose a tension between the power and the individual resistance it appeals to. Power as a behavioural conduit – *une conduite des conduites* – always implies a moment of resistance, namely the moment when individuals consider behavioural alternatives. Foucault sees power as the relation between individuals, when one steers the behaviour of the other, even though the other has the freedom to act differently. Power in this sense is a strategic situation that leads individuals to behave in ways to which they would not spontaneously commit themselves[37]. Resistance, Foucault writes, is always at the heart of the balance of power. And it is precisely at this elementary level that privacy comes in, since personal freedom embodies behavioural alternatives other than those induced by the power relation. In other words, privacy is the legal recognition of the resistance or reticence to behaviour steered or induced by power. From this point of view, privacy in a constitutional democratic state represents a legal weapon against the development of absolute balances of power, again proving privacy's essential role in such a state[38].

---

36 On the work of Robert Merton and others, see A. WESTIN, *Privacy and freedom*, London, The Bodley Head Ltd, 1970, 13 and 58. One could easily argue that persons who have not committed criminal acts should not fear being surveilled, but this argument is based on the erroneous assumption that a society or a group should or can function on the basis of the full observability of its members. The individual in virtually every society engages in this continuing process of what could be called *social distance setting*. It is one of the key universal dialectical processes in social life. The reason for the universality of this process is precisely that individuals have conflicting roles to play in any society. A certain degree observation will prevent members of the group from performing effectively. Some measure of leeway in conforming to role expectations is presupposed in all groups and makes social life possible. To have to meet the strict requirements of a role at all times, without some degree of deviation, is to experience insufficient allowances for individual differences in capacity and training and for situational exigencies which make strict conformity extremely difficult. The surveillance also constrains the observer, since he must decide whether or not to act against the non-complying person and must measure the effects of not acting on the group perception of authority.

37 Cf. M. FOUCAULT, "Deux essais sur le sujet et le pouvoir" in H. DREYFUS & P. RABINOW, *Michel Foucault. Un parcours philosophique*, Paris, Gallimard, 1984, 313-314: "L'exercice du pouvoir [...] est un ensemble d'actions sur des actions possibles: il opère sur le champ de possibilités où vient s'inscrire le comportement de sujets agissants: il incite, il induit, il facilite ou rend plus difficile, il élargit ou limite, il rend plus ou moins probable; à la limite il contraint ou empêche absolument; mais il est bien toujours une manière d'agir sur un ou sur des sujets agissants, et ce tant qu'ils agissent ou qu'ils sont susceptibles d'agir. Une action sur des actions".

38 So privacy imposes a balancing of power and resistance in all power relationships. And this does – or at least should – not only apply to the interference of the state. The list also includes the business sector, companies, trade unions, police, doctors, etc. The legal system gives examples – some successful, some not – of attempts

3.3 The non-absolute nature of privacy

Before going any further it is necessary to recall that affirming the essential role of privacy does not at all imply that privacy and the freedom it protects are absolute or inviolable values. On the contrary, notwithstanding privacy's core importance in a democratic constitutional state it is clear that it is a relatively weak fundamental right[39]. Actually, not a single aspect of privacy takes absolute precedence over other rights and interests. That includes confidentiality of the mail, physical integrity and control over personal information. Never does an individual have absolute control over an aspect of his/her privacy. If individuals do have the freedom to organise life as they please, this will only remain self-evident up to the point that it causes social or inter-subjective friction. At that stage, the rights, freedoms and interests of others, as well as the prerogatives of the authorities, come into play. The friction, tension areas and conflicts create the need for a careful balancing of the rights and interests that give privacy its meaning and relevance. This shows clearly that, although quintessential for a democratic constitutional state, because it refers to liberty, privacy is a relational, contextual and per se social notion which only acquires substance when it clashes with other private or public interests[40].

In general, the legal profession does not like to work with absolute values[41]. Of particular relevance to the privacy right is the republican argument against complete privacy. For the republican school absolute privacy may be detrimental to the building of citizenship. Rousseau, a classic example of the republican stand, stresses the need for the citizen to

to safeguard the privacy of individuals by protecting it against powerful interests. Police services cannot invade the privacy of a home at will. Welfare workers also have to operate within limits. Homeowners do not have the unlimited right, despite the absolute right to property, to check on their tenants. Employers cannot check on their personnel and their telecommunication exchanges at will. Banks and insurance companies are, in principle, limited in their freedom to gather, process and pass on personal information.

39 This is nicely illustrated by the fact that the ECHR *e.g.* recognises different sorts of human rights. The ECHR recognises some so called "hard core" or absolute rights that must be respected even in times of emergency when derogations to other rights are justified (art. 15 § 2 ECHR). Next to this there are "normal rights" (*e.g.* art. 5 and 6 ECHR) which can be derogated from in times of emergency (art. 15 § 1). Finally the ECHR counts four rights which can be legitimately restricted in terms of emergency but also under some specified conditions (art. 8-11 ECHR, the conditions for permissible restrictions are listed in the second paragraphs of these Articles). Privacy is one of these "restrictable rights".

40 In these cases (and on a case by case basis) it will be up to the legislator or the judge to determine how heavily privacy weighs against other rights and legitimate interests. But if privacy is found to prevail in a case, this will lead to a prohibition of interference.

41 C.R. SUNSTEIN, *Legal Reasoning and Political Conflict*, Oxford, O.U.P., 1996, p. 220.

participate in the public sphere to achieve "true freedom"[42], and warns that private concerns threaten the functioning of good government. The rise of self-interest would mean the end of the state[43]. A contemporary preference for public political engagement and suspicion of private action, is found in the work of Nancy Fraser. Her civic republicanism sees politics as people reasoning together to promote a common good that transcends the mere sum of individual preferences. The idea is that through deliberation members of the public can come to discover or create such a common good. In the process of their deliberations, participants are transformed from a collection of self-seeking, private individuals into a public-spirited collectivity, capable of acting together in the common interest. On this view, private interests have no proper place in the political public sphere[44]. Other contemporary defences of the virtues of public life are contained in Jürgen Habermas's *The Structural Transformation of the Public Sphere*[45] and Hannah Arendt's *The Human Condition*. Both authors were quoted above when discussing the need for privacy and opacity in a democratic state. However, both political thinkers are careful not to overestimate the importance of privacy. Relying heavily on Greek and Roman concepts, Arendt argues that, for people to be truly human, they need to live a more public life. For Arendt, politics is the vita activa or active life, where "human life in so far as it is actively engaged in doing something, is always rooted in a world of men and of manmade things which it never leaves or altogether transcends"[46]. She complains that the modern concept of privacy as sheltering us from social and political life is harmful as it does not allow us to become full humans, which can only

42 He holds that man acquires, together with civil society, moral freedom, which alone makes man the master of himself, "for to be governed by appetite alone is slavery, while obedience to a law one prescribes oneself is freedom". Cf. J. J. ROUSSEAU, *Du contrat social* (1762) in *Oeuvres complètes de Jean-Jaqcues Rousseau*, edited by B. GAGNEBIN and M. RAYMOND, Paris, Gallimard (Pléiade), 1964, Book 1, Chapter VIII, p. 365.

43 "As soon as public service ceases to be the chief business of the citizens and they would rather serve with their money than with their persons, the State is not far from its fall. [...] The better the constitution of a State is, the more do public affairs encroach on private in the minds of the citizens. Private affairs are even of much less importance, because the aggregate of the common happiness furnishes a greater proportion of that of each individual, so that there is less for him to seek in particular cares. In a well-ordered city every man flies to the assemblies; under a bad government no one cares to stir a step to get to them, because no one is interested in what happens there, because it is foreseen that the 'general will' will not prevail, and lastly because domestic cares are all-absorbing. Good laws lead to the making of better ones; bad ones bring about worse. As soon as any man says of the affairs of the State, 'What does it matter to me?', the State may be given up for lost"; J. J. ROUSSEAU, *o.c.*, Book 3, Chapter XV, p. 429.

44 N. FRASER, "Rethinking the Public Sphere" in C. CALHOUN (editor), *Habermas and the Public Sphere*, Cambridge Mass., MIT Press, 1992, 130; Ch. Hunter, *l.c.*, p. 3.

45 J. HABERMAS, *The Structural transformation of the public sphere.* Translated by Thomas Berger and Frederick Lawrence. Cambridge, Mass.: MIT Press., 1989.

46 H. ARENDT, *The Human Condition.* Chicago: University of Chicago Press, (1958), 1998, p. 22.

occur through engaging in an active public life with other people. "To live an entirely private life means above all to be deprived of things essential to a truly human life"[47]. Arendt comments that even the most intensely private actions cannot be fully understood until they are made public to others[48].

Opacity can also hide immoral and illegal actions. It can therefore be strongly argued that an ideal open society might be best served by total mutual transparency of all actors, as David Brin, treading in republican footsteps, contends[49]. However, there can exist good reasons and contexts to favour the elaboration of prohibitionary privacy/opacity rules, of which a number already do exist. They translate the need to shelter individuals against a too intrusive *conduite de la conduite* by more powerful social actors.

4 DATA PROTECTION AS A TOOL FOR TRANSPARENCY

4.1 Introduction

Since the 1970s, several European states have passed data protection legislation, that is, legislation protecting individuals from abuse resulting from the processing (i.e., the collection, use, storage, etc.) of personal data by public administrations and private actors. In general, these laws specify a series of rights for individuals and demand good data management practices on the part of the entities that process data ("data controllers"). The starting point of data protection is the desire to protect the citizen. Its purpose is to ensure that personal data are processed in ways that make it unlikely that personal integrity and privacy will be infringed or invaded[50].

It is impossible to summarise data protection in two or three lines. "Data protection" is a catch all term for a series of ideas with regard to the processing of personal data (cf. infra). Through the application of these ideas governments try to reconcile fundamental but conflicting values such

47 *Ibid.,* p. 58.

48 "Compared with the reality which comes from being seen and heard, even the greatest force of intimate life – the passions of the heart, the thoughts of the mind, the delights of the senses – lead an uncertain, shadowy kind of existence unless and until they are transformed, deprivatised and deindividualised, as it were, into a shape to fit them for public appearance" (*Ibid.,* 50).

49 D. BRIN, *The transparent society. Will technology force us to choose between privacy and freedom*, Perseus publ., 1999, p. 378.

50 P. BLUME, "The Citizens" Data Protection", *The Journal of Information, Law and Technology*, 1998/1 http://elj.warwick.ac.uk/jilt/infosoc/98_1blum/ (9p.), pp. 1-2.

as privacy, free flow of information, governmental need for surveillance and taxation, etc. In general data protection does not have the prohibitive nature of criminal law. The data subject does not own his data and he or she cannot in many cases prevent processing of his data. Under the current state of affairs, data controllers are recognised to have a right to process data pertaining to others. Hence, data protection is pragmatic in nature: it assumes that private and public actors need to be able to use personal information and that this in many cases must be accepted for societal reasons. The "thou shall not kill" that we know from criminal law, is replaced by a totally different message: "thou can process personal data under certain circumstances". The principled rupture in logic is evident.

4.2 The rationale behind data protection

Data protection is not prohibitive. On the contrary, in the public sphere, it is almost a natural presumption that public authorities can process personal data as this is necessary for the tasks they have to perform under statute, since, in principle, public authorities in democratic societies act on behalf of the citizens. The main aims of data protection consist in providing various specific procedural safeguards to protect individuals' privacy and in promoting accountability by government and private record-holders. Data protection laws were precisely enacted not to prohibit, but to channel power, viz. to promote meaningful public accountability, and provide data subjects with an opportunity to contest inaccurate or abusive record holding practices. The rationale behind data protection in the public sector is the knowledge that authorities can easily infringe privacy and that in all administrative systems there is an urge to collect, store and use data, an urge which must be curtailed by legal regulation. This is one of the functions of traditional administrative law and extends to data protection law[51]. A similar rationale explains the European option to regulate processing done in the private sector.

Data protection regulations mainly[52] belong to the tools of transparency, as opposed to the protection of privacy that pertains to the tools of opacity. The sheer wordings of the data protection principles (the

51 P. BLUME, "The Citizens" Data Protection", *l.c.*, pp. 2-3.

52 There are exceptions: namely those parts of the data protection regime that provide for a prohibition of processing (*e.g.* sensitive data, secretly collected personal data) actually fall under a privacy or opacity ruling. See *infra*.

fairness principle, the openness principle and the accountability principle, the individual participation principle, [...]) already suggest heavy reliance on notions of procedural justice rather than normative (or substantive) justice. The data protection regulations create a legal framework based upon the assumption that the processing of personal data is in principle allowed and legal[53]. As such these regulations implicitly accept that a processing of personal data is closely linked to the exercise of power and that it facilitates its establishment.

That explains why European data protection regulations were immediately conceived as applicable both in the public and in the private sector. The power of those, be it in the public or in the private sector, who process personal data concerning others (whether with the help of information technology or not) is generally already greater to begin with. The stream of personal data primarily flows from the weak actors to the strong. Citizens not only need to provide information to the authorities, but they also need to do so as tenant, job seeker, customer, loan applicant and patient. That is precisely why legal tools of transparency and accountability under the form of data protection regulations were devised for application both in the public and in the private sector.

4.3 Data protection as an opacity tool?

At first sight privacy and data protection are identical tools in nature, since the Data Protection Directive foresees a system of general prohibition, requiring some conditions to be met for "making data processing legitimate". The impression is given that the basic logic behind is of a prohibitive nature: "no processing, unless...". This understanding is however not correct for, firstly, the directive was heavily inspired by and had to accommodate existing national data protection regulations which were not based upon the prohibition principle. Secondly, the Data Protection Directive provides for a catch all ground for private data processing in its art 7 f. According to this article personal data can be processed without consent of the data subject if the processing "is necessary for the purposes of the legitimate interests pursued by private interests, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject".

53 An outright processing ban effectively applies only to special categories of sensitive personal data concerning racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, health or sex life.

For some authors this article even covers the processing of data for direct marketing purposes. Indeed such an article obliges a serious analyst to doubt and even refute the idea that the processing of personal data is in principle prohibited or dependent of the consent of the data subject. Art. 7 f in fact spans the whole scale of possibilities and can obviously "make data processing legitimate" for every thinkable business interest.

Nevertheless, exceptions from this general rule do exist. For instance, a prohibitive rule applies to "sensitive data" (data relating to racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership or data concerning health or sexual preference). The core of the underlying motive is that the processing of such sensitive data bears a supplementary risk of discrimination. The prohibition is nonetheless never absolute but derogations are (in principle) only possible in strictly defined circumstances, for example for reasons of national security. Another example can be found in Art. 15 Data Protection Directive[54], inasmuch as this article can be construed as the prohibition of decision making affecting persons solely on the basis of profiles. But again, both prohibitive features are accompanied by numerous exceptions that do not set strong limits to the targeted actions.

A third opacity tool in data protection can be found by an interpretation of the purpose specification principle. This principle, at the heart of data protection as it existed in many countries before the European Data Protection Directive came into force, states that the purposes for which personal data are collected should be legitimate and should be specified not later than at the time of data collection. Subsequent use should be limited to the fulfilment of those purposes. Preventive control of the intention of

54 According to this article every person has the right "not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data". The article refers to automated processing of data "intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.". The goal is to guarantee everyone's participation in important personal decisions. A dismissal based purely on the data from the company time clock is, as a result, unacceptable. It applies also to the rejection of a jobseeker based on the results of a computerised psycho-technical assessment test or to a computerised job application package. Those decisions have to take professional experience or the result of a job interview into account. The automated test is insufficient and it applies to such sectors as banking and insurance. The EU member states have to enact provisions that allow for the legal challenge of computerised decisions and which guarantee an individual's input in the decision-making procedures. However member states are allowed to grant exemptions on the ban on computerised individual decisions if such a decision "(a) is taken in the course of the entering into or performance of a contract, provided the request for the entering into or the performance of the contract, lodged by the data subject, has been satisfied or that there are suitable measures to safeguard his legitimate interests, such as arrangements allowing him to put his point of view; or (b) is authorised by a law which also lays down measures to safeguard the data subject's legitimate interests".

the controller and prohibition of illegitimate use was at the heart of data protection. The idea was not wholly new, since Article 8 ECHR holds that infringements of privacy can only be organised by law for legitimate purposes. In the past, when data was used mostly for only a single purpose, the logic behind the principle could be upheld without too much difficulty. Today, in the new economy and in a public sector ready for e-government, data is used for multiple purposes and much more intensely and effectively than ever before. Clearly these evolutions have influenced the drafting of the Data Protection Directive.

The fundamental purpose limitation principle is now stated in Article 6 (1b) of the Data Protection Directive. Compared to the situation in, *e.g.*, Belgium before the Data Protection Directive, the wordings are weakened. It is now said that subsequent use of data should be limited to the fulfilment of the initial purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose. This loosening of the purpose specification principle, coupled to the numerous possibilities in the Directive to render processing legitimate by obtaining consent, can be interpreted as a shift from prohibitive to channelling logic. Transparency seems to have replaced legitimacy as the core value of data protection. Whatever processing has been rendered transparent is legitimate.

4.4 The charter of fundamental rights of the European Union

The European Convention for the Protection of Human Rights (ECHR) plays a crucial role regarding the protection of privacy. However, in the ECHR there is no article that explicitly protects personal data. Although the organs of the ECHR have recalled on several occasions that data protection is an issue which falls within the scope of Article 8 ECHR[55], they have also held that not all aspects of the processing of personal data are protected by the Convention and that not all personal data are worthy of privacy protection[56]. In recent cases such as *Amann, Rotaru* and *P.G. and J.H. v. the United Kingdom*, the European Court seems to remedy to this by applying a

55 For instance: European Commission on Human Rights, *Lundvall v. Sweden*, 11 December 1985, case 10473/83, *D.R.*, v. 45, p. 130.

56 For a detailed discussion: P. DE HERT, "Mensenrechten en bescherming van persoonsgegevens. Overzicht en synthese van de Europese rechtspraak 1955-1997" [Human Rights and Data Protection. European Case-Law 1995-1997], in *Jaarboek* ICM 1997, Antwerpen, Maklu, 1998, pp. 40-96.

very broad definition of privacy[57]. However, these cases should be carefully interpreted. A closer reading shows that the old distinction between "data that merit protection" and "data that does not" is still operating. Also, the reference to existing data protection treaties is formulated in a way that leaves room for discretion.

Very recently the proper role of data protection has received constitutional recognition in Article 8 of the 2000 Charter of Fundamental Rights of the European Union[58]. Unlike in the ECHR the Charter did provide for a separate right to data protection next to the right to a private life for the individual. The Charter contains an uninspired copy of Article 8 ECHR, namely Article 7 which states that: "Everyone has the right to respect for his or her private and family life, home and communications". But the next and separate Article 8 of the Charter focuses explicitly on the protection of personal data. It inspiringly states:

> Everyone has the right to the protection of personal data concerning him or her. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data that has been collected concerning him or her, and the right to have it rectified. Compliance with these rules shall be subject to control by an independent authority.

This recognition of a constitutional right to data protection in the EU Charter should be welcomed for several reasons. To begin with it allows for a sensible constitutional division of labour. Data protection explicitly protects values that are not at the core of privacy, such as the requirement

57 For instance in *Amann v. Switzerland*, judgement of 16 February 2000, § 65-57: "The Court reiterates that the storing of data relating to the 'private life' of an individual falls within the application of Article 8 § 1 (see the *Leander v. Sweden* judgement of 26 March 1987, Series A n. 116, 22, § 48). It points out in this connection that the term 'private life' must not be interpreted restrictively. In particular, respect for private life comprises the right to establish and develop relationships with other human beings; there appears, furthermore, to be no reason in principle why this understanding of the notion of 'private life' should be taken to exclude activities of a professional or business nature (see the *Niemietz*, § 29 and *Halford v. United Kingdom,* judgement of 25 June 1997, § 42). That broad interpretation tallies with that of the Council of Europe's Convention of 28 January 1981 for the protection of individuals with regard to automatic processing of personal data, which came into force on 1 October 1985, whose purpose is 'to secure in the territory of each Party for every individual [...] respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him' (Article 1), such personal data being defined as 'any information relating to an identified or identifiable individual' (Article 2). In the present case the Court notes that a card was filled in on the applicant on which it was stated that he was a 'contact with the Russian embassy' and did 'business of various kinds with the company [A.]' (see paragraphs 15 and 18 above). The Court finds that those details undeniably amounted to data relating to the applicant's 'private life' and that, accordingly, Article 8 is applicable to this complaint also".

58 Http://europa.eu.int/comm/justice_home/unit/charte/en/charter02.html.

of fair processing, consent or legitimacy. The explicit recognition in the new provision of a "right of access to data that has been collected concerning him or her, and the right to have it rectified" solves legal problems unanswered by the case law of the European Court of Human Rights. Equally, there is no ground in this case law for a right to have compliance with (all) data protection rules controlled by an independent authority, as is foreseen by the last paragraph of the new provision[59]. Furthermore, the Charter extents the protection of personal data to private relations and the private sector[60].

The recognition of a separate right to data protection, next to privacy, is also more respectful of the different European constitutional traditions. Contrary to countries such as Belgium that have linked data protection from the start to privacy, countries such as France and Germany, lacking an explicit right to privacy in their constitution, have searched and found other legal anchors for the recognition of data protection rights. French data protection is therefore based on the right to liberty, whereas German data protection is based on the right to have human dignity recognised. Equally, the roots of American data protection are to be found in public law, viz. fair information practices. Hence, there are several grounds for data protection, which does not allow the use of one general privacy label.

Last and foremost, data protection has grown in response to problems generated by new technology. It brings no added value to reduce all these responses to "privacy". Other values and concerns are also at play. Take for instance the right not to be discriminated against that is protected by Article 15 of the European Data Protection Directive[61]. There is also a special regime for "sensitive data" in the Directive prohibiting processing of data relating to racial or ethnic origin, political opinions, religious or philosophical beliefs and so on. The connection with rights and liberties such as the freedom of religion, freedom of conscience and the political freedoms is obvious.

59 Article 13 ECHR (right to an effective legal remedy) does not create an independent right. The European Court refuses to consider issues under this provision, when there is no violation of another right of the ECHR. See *infra*.

60 Cf. Y. POULLET, "Pour une justification des articles 25 et 26 en matière de flux tranfrontières et de protection des données" in *Ceci n'est pas un juriste [...] mais un ami. Liber Amicorum Bart De Schutter*, M. COOLS, C. ELIAERTS, S. GUTWIRTH, T. JORIS & B. SPRUYT (editors), Brussels, VUBPress, 2003, p. 278.

61 According to this article every person has the right "not to be subject to a decision which produces legal effects concerning him or significantly affects him and which is based solely on automated processing of data". The article refers to automated processing of data "intended to evaluate certain personal aspects relating to him, such as his performance at work, creditworthiness, reliability, conduct, etc.". The goal is to guarantee everyone's participation in important personal decisions. A dismissal based purely on the data from the company time clock is, as a result, unacceptable.

5 THE SHIFT FROM OPACITY TOWARDS TRANSPARENCY IN EUROPEAN HUMAN RIGHTS LAW

5.1 European human rights law and the legality requirement

After having discussed the respective functions of privacy and data protection in the European democratic state, let us now turn to the European legal landscape of human rights law. In our introduction we mentioned the crucial role of the European Convention for the Protection of Human Rights regarding the protection of privacy. This Convention is designed to protect individuals' fundamental rights and freedoms and provides for a judicial procedure which allows individuals to bring actions against governments, if they consider that they are the victims of a violation of the Convention. After the exhaustion of national remedies, individual complainants have direct access to an international court, the European Court of Human Rights in Strasbourg.

Article 8 ECHR, the privacy article of the Convention, states:

> (1.) Everyone has the right to respect for his private and family life, his home and his correspondence. (2.) There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

This provision, partly copied in Article 7 of the European Union Charter of Fundamental Rights[62], is the source for EU legislation dealing with privacy and the protection of personal data, as well as of national legislation. Article 8 of the ECHR does not formulate privacy as an absolute right. Exceptions are made possible in the second paragraph of the provision, but the drafters of the Convention took care to provide safeguards against possible abuse of the right to formulate exceptions. Therefore, if any exception to the protection of data privacy is adopted respect has to be given to the conditions laid down in Article 8.2 of the Convention, that is, any invasion of privacy for a legitimate reason (for purposes of criminal investigation, usually the prevention of

62 This Article of the EU-Charter states that: "Everyone has the right to respect for his or her private and family life, home and communications".

crime) must be adopted "in accordance with the law" and when "necessary in a democratic society". Those requisites are cumulative[63].

Article 8 of the ECHR has been largely commented by legal authorities and applied by the European Court of Human Rights in Strasbourg[64]. We have analysed this case-law elsewhere[65], and limit ourselves to observations pertinent to the theme of this paper.

The large or extensive interpretation of the provision by the European Court is not limited to the scope of the notion of "privacy" included in Article 8.1 of the Convention[66]. The first requisite of Article 8.2 of the Convention, viz. a restriction must be adopted "in accordance with the law", has been interpreted in a non-formal way. In *Kruslin* the Court held, in the context of

63 "The interference was not therefore 'in accordance with the law' as required by the second paragraph of Article 8 and there has been a violation of this provision. In these circumstances an examination of the necessity of the interference is no longer required", *P. G. and J. H. v. The United Kingdom*, judgement of 25 September 2001, § 63. "The Court concludes that the interference cannot therefore be considered to have been 'in accordance with the law' since Swiss law does not indicate with sufficient clarity the scope and conditions of exercise of the authorities' discretionary power in the area under consideration. [...] Having regard to the foregoing conclusion, the Court does not consider it necessary to examine whether the other requirements of paragraph 2 of Article 8 were complied with"; *Amann,* § 19. See also: V. Coussirat-Coustere, "Article 8 § 2", in L. Pettiti, E. Decaux & P. Imbert (editors), *La Convention Européenne des Droits de l'Homme. Commentaire article par article*, Economica, 2e Edition, Paris, 1999, pp. 323-351.

64 The rights and freedoms of the Convention are formulated in a broad fashion. In order to apply them to concrete situations the Strasbourg authorities employ a number of interpretation techniques. Thus, the Court may embark on a grammatical analysis of the provision at issue, or apply a systematic interpretation; it may refer to the *travaux préparatoires* of the Convention or rather interpret the Convention according to "present-day conditions". An important element in this respect is also whether consistent State practice exists in regard to the issue at hand. One of the most remarkable aspects of the Strasbourg case-law is its dynamic character. On numerous occasions the Court emphasised that the Convention is "a living instrument which should be interpreted according to present-day conditions". See, *e.g.*, *Tyrer v. the United Kingdom*, judgement of 25 April 1978, § 31; *Marckx v. Belgium*, judgement of 13 June 1979, § 41; *Dudgeon v. the United Kingdom,* judgement of 22 October 1981, § 60; *Soering v. the United Kingdom*, judgement of 7 July 1989, § 102; *B. v. France*, judgement of 25 March 1992, §§ 45-48 and *Salesi v. Italy,* judgement of 26 February 1993, § 19. This culminated in 1995 when the Court held that the Convention "cannot be interpreted solely in accordance with the intentions of their authors as expressed more than forty years ago [...] at a time when a minority of the present Contracting Parties adopted the Convention" (*Loizidou v. Turkey (prel. obj.)*, judgement of 23 March 1995, § 71). In a similar vein the Court has repeatedly stressed that the Convention is intended to guarantee "not rights that are theoretical or illusory but practical and effective"; see, *e.g. Airey v. Ireland*, judgement of 9 September 1979, § 24 and *Soering*, § 87. It will be clear that this approach opens the way for expanding the protection offered by the Convention.

65 For a detailed analysis of the case-law, see P. DE HERT, *Artikel 8 EVRM en het Belgisch recht. De bescherming van privacy, gezin, woonst en communicatie* [Article 8 ECHR and the Law in Belgium. Protection of Privacy, House, Family and Correspondence], Gent, Mys en Breesch Uitgeverij, 1998, 367 p. and P. DE HERT, "Artikel 8 EVRM. Recht op privacy" [Article 8 of the Convention on Human Rights. The Right to Privacy] in J. VANDE LANOTTE & Y. HAECK (editors), *Handboek EVRM. Deel 2 Artikelsgewijze Commentaar*, Antwerp-Oxford, Intersentia, 2004, pp. 705-788.

66 *Rotaru v. Romania*, judgement of 4 May 2000; *P.G. and J. H.*, etc.

French criminal procedure, that the notion of "law" comprises written as well as unwritten law[67].

This questionable move towards flexibility (*infra*) has been partly compensated by the theory of the Court that the notion of "law" implies qualitative requirements, notably those of "accessibility" and "foreseeability". Interference by the executive with the rights and freedoms of the individual should not be permitted unless there is a clear legal basis to do so. It is not sufficient for Member States just to adopt a written law. According to the Court a legal and fully transparent legal basis must exist for justifying a measure limiting privacy rights. The object of the legal basis must also be specific. Any extensive exploratory or general surveillance (for example, of persons or of data) is prohibited[68]. A formal legal basis is not sufficient. The law, – be it case law or statute – must be of a certain quality: foreseeable (sufficiently detailed) and accessible and providing remedies for the citizen[69]. Partly these three quality requirements follow from the more general principle of the "rule of law". This principle requires effective safeguards against arbitrary interference by the authorities.

5.2 The success of the legality requirement

Many cases under Article 8 involve the legality requirement. Often the Court found that the safeguards needed to comply with the requirement were lacking[70]. While in continental "civil" legal systems and culture it is regarded as self evident that interference with the individual citizen by the state requires an explicit basis in law, common law systems take the opposite view: everything is allowed unless forbidden. Therefore, the full implementation of the ECHR required of the United Kingdom a substantial

67 See *Kruslin v. France*, judgement of 24 April 1990, §§ 27-36. At the same time the Court has realistically accepted that the wording of statutes is not always precise and that excessive rigidity must be avoided (*Kokkinakis v. Greece*, judgement of 25 May 1993, § 40; *S.W. v. UK*, judgement of 22 November 1995, § 36).

68 This principle has been constantly repeated by the European Court of Human Rights in what concerns electronic surveillance and wire tapping; see notably recently *Klass and others v. Germany*, judgement of 6 September 1978 and *Khan v. U.K.,* judgement of 12 May 2000.

69 The expression "in accordance with the law" requires, firstly, that the impugned measure should have some basis in domestic law; secondly, it refers to the quality of the law in question, requiring that it should be accessible to the person concerned, who must moreover be able to foresee its consequences for him, and that it is compatible with the rule of law. It also requires that the measure under examination comply with the requirements laid down by the domestic law providing for the interference. See *e.g. Kopp v. Switzerland*, judgement of 25 March 1998, § 55; *Perry v. United Kingdom,* judgement of 17 July 2003, § 55.

70 See, *e.g.*, *Kruslin*, §§ 30-35 and *Niemietz*, § 37.

cultural *volte face*[71], at least as far as the rules governing police powers are concerned[72]. A similar cultural or psychological revolution has taken place in the Civil law systems, such as France and Belgium, where historically a central role in law enforcement was accorded to the investigating judge, a person believed to be above all suspicion due to his independent position. The 1990 *Kruslin* and *Huvig* cases scattered the holy image of the investigating judge, unknown to Common Law systems. A provision in the French Code of Criminal Procedure giving this judge "all the necessary powers to seek the truth" was declared incompatible with Article 8 of the Convention. Even the powers of the investigative judge have to be made foreseeable and accessible[73].

The case-law regarding the legality requirement has also enabled the European Court to make clear that assessing privacy infringements is not only a question of transparency, but also of accountability. The requirement of foreseeability in particular has allowed the Court to fuse the due process requirements of Article 6 ECHR and the effective remedy requirement of Article 13[74] into the privacy requirements of Article 8 ECHR. With regard to telephone tapping and other investigation techniques this has brought the Court to a very detailed set of conditions that have to be fulfilled by the legislatures and have to be respected by law enforcement authorities. These conditions oblige legislators to explicitly and precisely foresee which categories of persons can be the object of the measures, for which incriminations the measures can be taken, how long they can last, how reports/log books about the measures must be made up and, in case of later suspension of prosecution or acquittal, be destroyed. However, this has at least one important drawback: any measure of telephone tapping will be deemed to be legitimate if it meets the conditions. It can thus also be feared that such a channelling "transparency" approach encroaches upon the normative "opacity" that privacy is meant to protect.

---

71 The course of the events from *Malone* to the Regulation of Investigatory Powers Act 2000 is a *locus classicus*. A violation was found of Article 8 ECHR in a case concerning interception and metering of telecommunication, on the grounds that a legal basis as required by Article 8 was absent in English law. The Interception of Communications Act 1985 was an attempt to remedy for this. Cf. *Malone v. United Kingdom*, judgement of 2 August 1984.

72 P. ALLDRIGDE & CH. BRANTS, "Introduction" in P. ALLDRIGDE & CH. BRANTS (editors), *Personal Autonomy, the Private Sphere and the Criminal Law. A Comparative Study*, Oxford, Hart Publishing, 2001, p. 13.

73 P. DE HERT, "Het recht op een onderzoeksrechter in Belgisch en Europees perspectief", *l.c.*, pp. 155-198.

74 This article reads as follows: "Everyone whose rights and freedoms as set forth in [the] Convention are violated shall have an effective remedy before a national authority [...]".

5.3 A critical comment about the Strasbourg focus on the legality requirement

Undoubtedly the legality requirement has important merits and by imposing it for police powers (inter alia) that touch upon certain human rights its basic goal is realised. Although we suggested in our introduction that privacy is not a central value of criminal law enforcement, the aforementioned case-law shows that privacy has an impact on criminal law and has the ability to enrich it[75]. Also, the way the European Court handles this requirement is not without intelligence and constitutional wit. But strictly speaking the legality requirement in Article 8 ECHR has nothing to do with privacy[76]. Privacy is not about legality, it is about power and how to stop it. This brings us to the ambiguity we mentioned in our introduction when referring to the vast body of privacy case-law developed by the European Court of Human Rights in Strasbourg and also to a critical attitude regarding the dominant privacy focus on the legality requirement laid down in Article 8 of the Convention[77].

As an opacity tool Article 8 of the ECHR is not an ideal starting position, especially when compared to *e.g.* the U.S. First Amendment that foresees no explicit exceptions. To start with, the right to respect for privacy as enshrined in the Convention is not absolute. The flexible notion of respect is informed by the interests of national security, public safety, the economic well being of the country, prevention of disorder and crime, protection of public morals and the rights and freedom of others. This broadly formulated list of legitimate grounds to restrict privacy in Article 8.2 of the Convention potentially allows for a broad governmental discretion. The nature of the privacy right enshrined in the Convention is therefore not clear from the start. Its use and interpretation decide whether and when this provision functions as a power blocking tool. The responsibility of the European Court, as a last interpreter of the Convention, is considerable in this respect.

---

75 Generally speaking there is no legality requirement within the law of criminal procedure. Although we observed that in continental "civil" legal systems and culture it is regarded as self evident that interference with the individual citizen by the state requires an explicit basis in law, almost no countries exist, with the exception of Germany and the Netherlands, where the Codes of Criminal Procedure state that all actions of the law enforcement authorities need an explicit basis in law. Also in human rights law, especially with regard to the law of criminal procedure, there is no "human" right to have all police powers based upon a law. A general requirement for legality with regard to all law enforcement powers is, for instance absent in the ECHR. By saying that certain powers touch upon privacy a bit of this "principle" can however be read into the Convention.

76 The requirement can also be found in the three subsequent articles of the Convention.

77 We borrow from P. DE HERT, "Strafrecht en privacy. Op zoek naar een tweede adem" [Criminal Law and Privacy. Searching for a New Breath], *Rechtshulp. Maandblad voor de sociale praktijk*, 2003/10, pp. 41-54.

We argued above that data protection regulations should be regarded as transparency tools, whereas the protection of privacy pertains to the tools of opacity. It is useful and necessary not to blur this distinction since each tool has it proper logic. However, the focus on the requirement "in accordance with the law" has turned Article 8 ECHR into a transparency-promoting vehicle. One can regret, but understand, the habit of the Court not to go any further with its investigations once it has decided a case merely on the basis of the requirement "in accordance with the law"[78]. However, one cannot understand its insistence on the quality of the law in the context of Article 8 of the Convention and its disregard of the formal status of the legal basis that is used by the Member States to justify certain privacy limitations. We recall that even within the Civil Law systems it is not always required to have a formally voted law. Case-law and other legal texts may do, as long as they fulfil the quality requirements. Democratically there is a loss. Insisting on the content of law, rather than on the formal basis of law, has allowed the Court to declare certain regulations "in accordance with the law" that have not even been debated and approved by a parliament[79].

5.4 The danger of proceduralisation

Constitutionally there is also loss, because the proper division of labour between the existing human rights has not been respected[80]. We have already said that the requirement of foreseeability has allowed the Court to fuse the due process requirements of Article 6 ECHR and the effective remedy requirement of Article 13 into Article 8 ECHR. This approach has been detrimental for both provisions, especially Article 13. Elements of

78 The other requirements of Article 8.2 of the Convention, especially the requirement that privacy limitations need to "necessary in a democratic society", are often not checked by the Court, when a breach of the legality requirement is found. The Convention organs treat the requirements as successive hurdles. This means that where they find that a measure complained of is not "in accordance with the law", then they do not proceed to examine whether the measure satisfies the requirement of "necessity in a democratic society". For instance in *P.G. and J.H.*, § 38: "As there was no domestic law regulating the use of covert listening devices at the relevant time [...], the interference in this case was not 'in accordance with the law' as required by Article 8 § 2 of the Convention, and there has therefore been a violation of Article 8 in this regard. In the light of this conclusion, the Court is not required to determine whether the interference was, at the same time, 'necessary in a democratic society' for one of the aims enumerated in paragraph 2 of Article 8." See also: I. CAMERON, *National Security And The European Convention On Human Rights*, The Hague/London/Boston, Kluwer Law International, 2000, p. 36.

79 See: P. BLONTROCK & P. DE HERT, "Telefoontap: Tournet, Peureux, Hüvig, Kruslin et les autres", *Rechtskundig Weekblad*, 1991-1992, 865-871 and I. Cameron, *o.c.*, p. 34.

80 The ECHR recognises some procedural rights, such as the right to a fair trial (art 6) and the right to an effective remedy (art. 13) next to substantial rights like privacy, freedom of expression, conscience, religion and assembly (art. 8 to 11).

procedural rights are borrowed to construe a substantive norm, and the result is used to interpret the procedural rights narrowly[81].

Of course, this situation is far from evident. The two categories of rights in the Convention, procedural rights and substantive rights, were originally conceived as being complementary, but today it can be held that, in the case law of the Court, the two categories have merged or "conpenetrated"[82]. More and more the respect of a number of procedural conditions concerning transparency, impartiality, accessibility, [...] has become an essential condition for a legitimate restriction of a substantial conventional right. According to Tulkens and Van Droogenbroeck this "proceduralisation of substantial rights" has both positive and negative effects. It is positive inasmuch as this evolution contributes to the objectivity and credibility of the Court's control as it compels the Court to take "distance" from the facts of the case: prior to an inquiry into the merits and facts of the case (entailing a discussion of the state's *margin of appreciation*), the Court will merely check if the contested decision or measure meets all formal requirements. Indeed, such a demarche also meets the principle of the subsidiarity of the international review. On the other hand, negatively speaking now, Tulkens and Van Drooghenbroeck fear that a procedural review might come to be substituted for the substantial review by a procedural one: the proceduralisation of substantial rights might well lead to a situation wherein the Court would definitively stop to check and review restrictions of fundamental rights in the light of the values and norms enshrined in the ECHR[83].

If proceduralisation bears the advantage of objectivity and impartiality, it also leads just as surely to the formalisation, bureaucratisation and depoliticisation of human rights questions: meeting formal constraints and conditions is never a hurdle too high to take. This is convincingly illustrated

81 In *Klass* the Court introduced a heavily criticised restrictive or "relative" approach. The Court considered that Article 13 has a subsidiary character in relation to Article 8. Thus Article 13 could not be interpreted so as to nullify the efficacy of the measures of secret surveillance already found to be compatible with Article 8. The Court stated that "an effective remedy [...] must mean a remedy which is as effective as can be having regard to the restrictive scope for recourse inherent in any system of secret surveillance"; *Klass*, § 69. See the consequences of this approach for judicial checking on secret surveillance: I. Cameron, *o.c.*, pp. 36-39.

82 In this context Françoise Tulkens and Sébastien Van Droogenbroeck use the term of "*conpénétration*". They come to the conclusion that "chacune des dispositions conventionnelles consacrant un droit substantiel est susceptible de sécréter des garanties d'ordre procédural contribuant à l'effectivité du droit concerné et attachées davantage aux processus décisionnels qu'aux décisions proprement dites"; FR. TULKENS & S. VAN DROOGENBROECK, "La cour européenne des droits de l'homme depuis 1980. Bilan et orientations" in *En toch beweegt het recht*, W. DEBEUCKELAERE & D. VOORHOOF (editors), *Tegenspraak-cahier 23*, Brugge, Die Keure, 2003, p. 224.

83 FR. TULKENS & S. VAN DROOGENBROECK S., *l.c.*, pp. 223-227.

by the telephone tapping case-law (*Klass, Malone, Huvig, Kruslin*, [...]) which has led to the devising of a detailed and elaborate set of conditions that taps must fulfil in order to be considered as legitimate. But the other side of the coin is that the question of the necessity of such practices in a democratic society fades away into the background: it is enough that the tapping meets all the formal conditions [...] Proceduralisation might well bring the erosion of recognised rights[84].

If there are good reasons for this process of proceduralisation, the Court has not provided them[85]. The unfortunate treatment of Articles 6 and 13 in *Klass* seem to be inspired by the will to accept secrecy and to strain the paradigmatically adversarial nature of the judicial guarantees set forth in Article 6[86]. The attempt to read Article 6 notions in the legality requirement

84 Of course, we are well aware that emphasising the substance of the fundamental rights also has dangerous drawbacks. Indeed, such an approach implies that Court determines and describes what these rights consist in, which is only possible through the identification of the values upon which the ECHR relies. But which are these values? What is the meaning of words like "democracy" and "rule of law" setting the broader environment in which the "fundamental rights and freedoms" must be concretised? Which political philosophy lies at the roots of the Convention: is it the aim to implement a minimal and liberal state wherein individual liberty is the most important value? Or does the Convention aim at the realisation of a more republican or even communitarian state driven by the public interest or common values? Or is it something between the two: a kind of relational or polyphonic state?

85 It is possible to find some rationale to transform Article 8 into a procedural norm. In *Silver and Others v. United Kingdom* the Court made it clear that a law which "allows the exercise of unrestrained discretion in individual cases will not possess the essential characteristics of foreseeability and thus will not be a law for present purposes. The scope of the discretion must be indicated with reasonable certainty"; *Silver and Others v. United Kingdom*, judgement of 25 March 1983, § 88-89. In cases such as *Klass, Huvig* and *Kruslin* the Court has also stated that adequate safeguards also must exist against abuse of the discretion established by law (*Klass*, § 63*, Huvig v. France*, judgement of 24 April 1990, § 34 and *Kruslin,* § 35).

86 In *Klass* it accepted that a secret telephone tapping measure was compatible with article 8 ECHR. One step further it decided that a legitimate *and* secret telephone tap cannot violate articles 6 and 13 ECHR. In their submission the applicants argued that the legislation violated Article 6 insofar as it did not require notification to the person concerned in all cases after the termination of surveillance measures and excluded recourse to the courts to test the lawfulness of such measures. The Court simply refused to consider the complaint: once decided that a system of secret surveillance without notification does not contravene Article 8, the right to judicial control in Article 6 does not apply (cf. *Klass*, § 75). But these articles do not provide for the same restrictions and exceptions as article 8, which permits us to conclude with François Rigaux that'(e) n développant son raisonnement à partir de l'équilibre entre la règle et l'exception dans l'article 8 la Cour introduit la même exception dans les articles 6 et 13 qui ne la connaissent pas". Cf. F. RIGAUX, *La protection de la vie privée et des autres biens de la personnalité*, Brussel/Parijs, Bruylant/L. G. D. J., 1990, 201. See also: E. A. ALKEMA, "Klass e.a. tegen de Duitse Bondsrepubliek", *Ars Aequi*, 1979, 323 and P. DE HERT, *Art. 8 E.V.R.M. en het Belgisch recht, o.c.*, 342-343. This is even more disturbing because the effect of articles 6 and 13 upon secret investigative methods should not be underestimated. Both articles foresee the right to have conflicts and disputes reviewed by an independent and impartial tribunal. This would imply that persons who were subjected to secret measures (and hence did not know about it) should afterwards be informed of these measures even if they remained without results. See also A. H. J. SWART, "Anoniem gerechtelijk vooronderzoek", *Ars Aequi*, 1984, 335. If this guarantee is lacking, only these subjects of the said secret measure who will be prosecuted later will be informed about the way data about them has been collected. Comp. "Au contraire, si les éléments recueillis sont insuffissants pour que des poursuites puissent être intentées ou que l'autorité compétente estime cellesci inopportunes, les organes du pouvoir exécutif ont le pouvoir discrétionnaire de conserver le secret sur les mesures prises à l'égard de faits qui, par hypothèse, n'ont entraîné aucune poursuite alors que, seul, leur caractère punissable justifiait que la mesure fût prise";

of Article 8 seems to be the result of a more or less explicit refusal by the Court to test the reasonableness of a system of secret surveillance provided for by law and in particular to test the requirement of subsidiarity. The citizen gets procedural guarantees as a compensation for the lack of testing of the reasonableness of the intrusion[87]. We recall that this case law misreads the convention by fusing together two (or three) provisions, which are motivated by very different ideas and do not run into each other. Article 8, unlike Article 6, has a general scope, while Article 6 is limited to conflicts about 'the determination of civil rights and obligations or of any criminal charge against" a person. Article 8 is about substantive issues, Article 6 about procedural rights. Unlike the fair trial clause, the privacy right speaks not only to procedural fairness but (mainly) to substantive fairness. It is therefore far more sensible to try to read the privacy right, with its open-ended nature, in light of other *substantive* norms contained in the Convention.

Ian Cameron is obviously of the same opinion. This author holds that these procedural requirements should not be seen as a feature of the legality requirement, but rather as a requirement of Article 13 and/or a requirement

F. RIGAUX, *o.c.*, 200. This looks a little bit like a carte blanche for secret measures. In sum, in *Klass*, the interpretation of the substantial right to privacy leads to a weakening of the procedural guarantees of the fair trial. For Rigaux *Klass* is a "jugement d'opportunité politique". In the Malone case also, the Court avoided reviewing a telephone tapping measure from the perspective of art. 13 arguing that this was unnecessary because it had already decided a violation of art. 8 ECHR. Recently, in *Perry* the Court sharply separated the privacy right from the fair trial right. *Perry*, § 48. "Though the Government has argued that it was the quality of the law that was important and that the trial judge ruled that it was not unfair for the videotape to be used in the trial, the Court would note that the safeguards relied on by the Government as demonstrating the requisite statutory protection were, in the circumstances, flouted by the police. Issues relating to the fairness of the use of the evidence in the trial must also be distinguished from the question of lawfulness of the interference with private life and are relevant rather to Article 6 than to Article 8. It recalls in this context its decision on admissibility of 26 September 2002 in which it rejected the applicant's complaints under Article 6, observing that the obtaining of the film in this case was a matter which called into play the Contracting State's responsibility under Article 8 to secure the right to respect for private life in due form".

87 See *Klass*, § 49 & 50: "49. As concerns the fixing of the conditions under which the system of surveillance is to be operated, the Court points out that the domestic legislature enjoys a certain discretion. *It is certainly not for the Court to substitute for the assessment of the national authorities any other assessment of what might be the best policy in this field* (cf., mutatis mutandis, the *De Wilde, Ooms and Versyp* judgement of 18 June 1971, § 93, and the *Golder* judgement of 21 February 1975, § 45; cf., for Article 10 § 2, the *Engel and others* judgement of 8 June 1976, § 100, and the *Handyside* judgement of 7 December 1976, § 48). Nevertheless, the Court stresses that this does not mean that the Contracting States enjoy an unlimited discretion to subject persons within their jurisdiction to secret surveillance. The Court, being aware of the danger such a law poses of undermining or even destroying democracy on the ground of defending it, affirms that the Contracting States may not, in the name of the struggle against espionage and terrorism, adopt whatever measures they deem appropriate. 50. *The Court must be satisfied that, whatever system of surveillance is adopted, there exist adequate and effective guarantees against abuse.* This assessment has only a relative character: it depends on all the circumstances of the case, such as the nature, scope and duration of the possible measures, the grounds required for ordering such measures, the authorities competent to permit, carry out and supervise such measures, and the kind of remedy provided by the national law" (italics are added). Comp. these paragraphs with §76 of the *Peck* judgement; *Peck v. United Kingdom,* judgement of 28 January 2003.

of "necessity in a democratic society"[88]. We prefer the first option. Article 8 of the Convention is no place for procedural questions. The framers of the Convention have designed other articles for that purpose. The transformation of Article 8 into a source of procedural rights and procedural conditions takes it away from the job it was designed for, viz. to prohibit unreasonable exercises of power and to create zones of opacity.

5.5 A requirement fundamental to opacity: necessary in a democratic state

We observed that a lot of important Article 8 cases are "solved" by concentrating on issues concerning transparency. The question whether a certain practice is "necessary in a democratic society" is often not answered by the Court, when a breach of the legality requirement is found (*supra*). However, checking on the legality requirement is of a fundamentally different nature than checking on the requirement "necessary in a democratic society". Only the latter requirement deals with the political question whether power should be limited, stopped or prohibited or, in other words, whether "opacity" must be protected. Even if a restriction of privacy is foreseen by law and serves an objective summed up in article 8 § 2 ECHR, this restriction must still be "necessary in a democratic society" and shouldn't reach further than what is strictly necessary[89]. This condition inevitably implies an ultimate balancing of interests, a value judgement and/or a substantial choice, which cannot be found in an exegetic reading of the text, nor in a strict application of logical rules[90].

Such a balancing of interests, which takes the weight of fundamental rights and freedoms duly into account, is essential[91]. It allows for the exercise of the political function of human rights. The need to limit power through opacity tools explains the importance of the requirement "necessary in a democratic society". Behind this requirement lies the true constitutional question with regard to law enforcement and privacy: is there a justifiable necessity for the law enforcers to break into the privacy and to lift the

88 I. Cameron, *o.c.*, 34.

89 About this condition see K. RIMANQUE, "Noodzakelijkheid in een democratische samenleving – een begrenzing van beperkingen aan grondrechten", in *Liber Amicorum Frédéric Dumon,* Antwerp, Kluwer Rechtswetenschappen, 1983, deel II, p. 1220.

90 K. RIMANQUE, *l.c.*, p. 1229.

91 Cf. S. GUTWIRTH, "De toepassing van het finaliteitbeginsel van de Privacywet van 8 december 1992 tot bescherming van de persoonlijke levenssfeer ten opzichte van de verwerking van persoonsgegevens" [The application of the purpose specification principle in the Belgian data protection act of 8 December 1992], *Tijdschrift voor Privaatrecht*, 4/1993, pp. 1409-1477.

opacity of the individual? In the context of Article 10 ECHR (freedom of expression) the Court has observed that "necessary [...] is not synonymous with *indispensable,* neither has it the flexibility of such expressions as *admissible, ordinary, useful, reasonable* or *desirable,* but that it implies a *pressing social need"*[92].

Obviously this interpretation is too far reaching for the European judges as regards privacy. Not many cases under Article 8 repeat this kind of exercises. Almost always the requirement of "necessity" is brought back to the question of proportionality, in some cases supplemented by the requirement that the reasons for the interference are relevant and sufficient[93]. Only in the recent *Peck* judgement can one find some word games referring to the semantic exercise in the context of Article 10 discussed above[94]. What is "proportionate" will depend on the circumstances. According to M. Delmas-Marty, in determining proportionality the Court takes particularly into account the nature of the measure taken (its reach, whether it is general of absolute, its adverse consequences, the scope for abuse of the measure), whether the state concerned could have taken other measures or implemented them in a less drastic way, the status of the persons involved whose rights can legitimately be subject to greater limitation (*e.g.* prisoners) and finally, whether there are any safeguards which can compensate for the infringement of rights which a measure can create[95].

However, the Strasbourg judges are too hesitant and reluctant to really address these issues. They clearly prefer the much more secure test of

92 *Handyside v. United Kingdom*, judgement of 7 December 1976, § 48.

93 P. DE HERT, *Artikel 8 EVRM en het Belgisch recht*, *o.c.*, 40-60. Compare with §76 of the *Peck* judgement: "In determining whether the disclosure was "necessary in a democratic society", the Court will consider whether, in the light of the case as a whole, the reasons adduced to justify the disclosure were "relevant and sufficient" and whether the measures were proportionate to the legitimate aims pursued"; *Peck v. United Kingdom,* judgement of 28 January 2003.

94 See the use of the term "pressing social need" in the following quote: "In such circumstances, the Court considered it clear that, even assuming that the essential complaints of *Smith and Grady* before this Court were before and considered by the domestic courts, the threshold at which those domestic courts could find the impugned policy to be irrational had been placed so high that it effectively excluded any consideration by the domestic courts of the question of whether the interference with the applicants" rights answered a pressing social need or was proportionate to the national security and public order aims pursued, principles which lay at the heart of the Court's analysis of complaints under Article 8 of the Convention; *Peck*, § 100.

95 M. DELMAS-MARTY, *The European Convention for the Protection of Human Rights,* Dordrecht, 1992, 71 quoted by I. Cameron, *o.c.*, 26. About proportionality see also: S. Gutwirth, "De toepassing van het finaliteitbeginsel [...] [The application of the purpose specification principle ...]", *l.c.*, § 20; S. Van Drooghenbroeck, *La proportionnalité dans le droit de la convention européenne des droit de l'homme. Prendre l'idée simple au sérieux*, Bruxelles, Bruylant/Publications des FUSL, 2002, 790 p. and W. Van Gerven, "Principe de proportionnalité, abus de droit et droits fondamentaux", *Journal des Tribunaux*, 1992, pp. 305-309.

the legality requirement (is there a law?)[96]. When they *have* to consider the "necessity" requirement, they replace it by the more flexible proportionality test. This is regrettable because there is a structural lack of information about the notion of the democratic state, which should be filled out precisely by the legislature and the judges. Assertions such as "we must defend our democracy against terrorism" give the impression that democracy is a kind of substance existing in itself. We never get to know what this substance is, but we are permanently invited and tempted to accept the proposition that we have to defend this democracy by all means, even undemocratic ones[97]. This is dangerous. Democracy is no substance but a whole of practices and processes that we associate with the label "democratic". It consists of a permanent questioning of all the practices and processes which make a society democratic. The core of constitutional thinking is to participate in that questioning by focussing upon some values that are considered to be fundamental.

6 COMBINING PRIVACY AND DATA PROTECTION

6.1 Combining the tools

In the preceding pages we critically discussed the shift from opacity towards transparency in European human rights law. There are numerous ways to revitalise or to protect the proper function of privacy and to make it relevant again for framing law enforcement.

On the level of case law, privacy could regain something of its esteem if the judges of the European Court of Human Rights would stop focusing only upon the legality criterion (if a statute makes everything possible, human rights are threatened and not protected) and start again taking the "necessary in a democratic society" test seriously: judges in a constitutional court have a crucial political role to play in drawing the line between acceptable and non-acceptable power, setting the switch between opacity tools and transparency tools.

96 Cf. I. Cameron, *o.c.*, p. 35.

97 J. BLOMMAERT, "Veiligheid en democratie?" [Security or Democracy?], in Liga voor Mensenrechten (editor), *Wordt de Europese ruimte van vrijheid, veiligheid en rechtvaardigheid een politiestaat?*, Gent, Liga voor Mensenrechten, 2003, (67-70), 67.

Maybe one day a legal culture will be firmly established in which it is regarded as self evident that interference with the individual citizen by the state requires an explicit basis in law, even in common law systems that have taken the opposite view in the past[98]. Hence, one can expect a future rich with cases in which the focus is on the requirement of "necessity in a democratic society"[99].

On the level of the legislator more attention should be paid to the distinct nature of the two sorts of legal tools that were invented to cope with power in a democratic constitutional state: the opacity tools establishing the limits of interferences in an individual's life by public and private powers on the one hand, and the transparency tools channelling and regulating the legitimate use of power on the other hand. Ideally, every time the legislator acted he would consider both tools and identify the kind of tool necessary for a given problem. How much of which tool is necessary when? Each tool supplements and pre-suppose the other. Channelling power in the mist is doomed to fail; limits and points of departure are necessary. Approaching new phenomena and new possibilities for law enforcement with heavy prohibitions may circumvent the legitimate interest of the state or block potentially interesting developments *e.g.* with regard to the use of new technology[100]. It may also lead to a situation in which the prohibitions are not respected. This would leave power relations uncontrolled, due to the lack of tools. Hence, an approach based mainly on opacity tools should be considered with due care.

If no specific action is taken towards new police or private demands to use new technological means, in a European context, data protection will apply. This would however imply, due to the enabling logic of this legal framework, that we say "yes" to all new forms of power applications,

98 P. ALLDRIGDE & CH. BRANTS, *l.c.*, 13.

99 Case law such as the *Khan* judgement, refusing to apply the principle that illegaly obtained privacy evidence should be rejected, seems to contradict this view of the future. By weakening the importance of privacy, privacy will lose its capacity as an opacity tool. The *Khan* doctrine (followed in cases such as *Doerga v. the Netherlands* and *P. G. and J. H. v. The United Kingdom*) is discussed in P. DE HERT, "De soevereiniteit van de mensenrechten: aantasting door de uitlevering en het bewijsrecht" [Sovereignity of human rights: threats created by the law of extradition and by the law of evidence], *Panopticon*, *Tijdschrift voor strafrecht, criminologie en forensisch welzijnswerk*, 2004, v. 25, n. 3, 229-238 and in P. DE HERT & F. P. ÖLCER, "Het onschadelijk gemaakte Europees privacybegrip. Implicaties voor de Nederlandse strafrechtspleging" [The notion of privacy made innocent. Implications for criminal procedure], *Strafblad. Het nieuwe tijdschrift voor strafrecht*, 2004, v. 2, n. 2, pp. 115-134

100 This approach is followed *e.g.* in Article 13 of the Charter of fundamental rights of European Union of 7 December 2000 prohibiting "eugenic practices" in particular those aiming at the selection of persons and in "making the human body and its parts a source of financial gain".

even when their legitimate character can be disputed. Also, data protection legislation tends to be very difficult and technical. This may give way to erosion and a denial of this new area of law. The strength of data protection, however, is not to be neglected. *The complex question "is this a privacy issue?" is replaced by a more neutral and objective question "are personal data processed?".* Data protection, as such, is a general framework for all kinds of surveillance: the written word, sounds, images, DNA and even smells can be understood as personal data falling under the scope of data protection. For instance, in the case of closed circuit TV (CCTV) – a technology that can be legally installed and operated, but very easily abused and used for unlawful purposes (for example by turning the camera away from the machines and towards the working personnel) – the attention of data protection for all processing aspects (from collection to destruction) is a guarantee. Moreover, data protection brings the issue of consent and the waiver of rights into focus, by making it explicitly possible in some cases (but not all), and only under specific conditions. We don't see how privacy case law could form an alternative for the elaborate list of rights and duties foreseen by the data protection bills. (How many cases and how many precedents are needed to achieve the same results?).

Furthermore, it should be stressed that the two approaches do not exclude each other. They depend on policy choices, which can be revised and adapted. As a result, pursuit of the transparency approach (regulating instead of prohibiting) might after some time and practice show that the opacity approach is preferable (or vice versa) or that a better balance between the two approaches should be devised. In reality one will rarely find legal solutions based exclusively upon one tool. A blend of the two approaches will generally be preferable, since a solid legal framework should be both flexible (transparency) and firmly anchored in intelligible normative choices (opacity). A good example of such balancing of approaches is given in Directive 2002/58/EC on privacy and electronic communications of 12 July 2002 (supra). This Directive puts an end to the long-lasting controversy regarding direct marketing by explicitly adopting an opt-in system which inherently implies the prohibition of unsolicited marketing mail unless the user makes an explicit request to receive it. In this example it becomes clear how the model of channelling business practices (transparency) is supplemented by the limiting model of a negative obligation (opacity) after due consideration and debate. A second example can be found in national legislations dealing with CCTV, containing for instance prohibitions on

directing cameras towards the entrances of private premises. A third example is provided by the numerous national bills on the use of DNA--samples in criminal matters. Although the processing of DNA-samples, from the perspective of data protection (Directive no. 95/46/EC), is in fact an ordinary application of processing of personal data, the riskiness of the matter explains why states supplement general data protection bills with specific prohibitive bills on DNA.

Of course our approach raises the question of the criteria for the application of the distinct tools to the (new) developments and techniques. When will opacity (privacy) be called upon, when will transparency (data protection) apply? Such criteria and norms should be used to set the switch between a privacy-opacity approach (prohibitive rules that limit power) and a data protection-transparency approach (regulations that channel power). Especially when faced with new problems, such as the insistence on security (of various government initiatives) or the development of new technologies, the approach should consist in combining the tools appropriately. The issue of the criteria is crucial because it will determine the sort of legal measures that will be taken and applied. The differences are sensible. On the one hand, legal measures from the privacy/opacity perspective include the elaboration (and strengthening) of prohibitionary legal regimes and protect and enforce the anonymity of behaviour (*e.g.* through approving regulations of techniques of anonymity, pseudonymity, identity management, [...]). On the other hand, legal measures from the data protection/transparency perspective are directed towards the data controllers allowing "to watch the watchdogs".

With regard to new technologies, the (European) legislator will have to assess the risks and threats to individual liberty separately[101]. The two complementary instruments at his disposal allow for a well-balanced regulatory framework. It can be assumed that there will be reliance on data protection and other transparency tools by default and that only in rare cases or after due consideration of actual risks will prohibitive opacity measures be taken to protect rights and freedoms and to promote trust in the Information Society. The sheer fact that both instruments co-exist implies a permanent determination to assess the level of acceptance and implementation of use

101 Anyhow, in dealing with future technologies with still unknown potential and bearing risks for the liberty of the individual, we should adopt a *precautionary* approach (a process that includes information gathering, broad consultation, participative procedures of decision-making, etc.).

and potential abuse of new technologies and the ensuing enforcement of legal rules. This process may explain why factors such as September 11 and new technological developments can account for a shift from transparency tools to opacity tools (when trust is fragile) or vice versa (when trust is reestablished).

6.2 Determining the switch

The importance of the issue of determining the criteria for the handling of the switch leading to a policy of opacity or transparency cannot be underestimated. Of course the choices concerning these criteria will be made by the legislators, preferably after a large and informed societal debate amongst the totality of those concerned (business, government, citizens, civil liberties groups, privacy advocates, [...]) and by competent judges at national and international level. For our part, we can only deal with the question of the criteria from a prospective, exploratory and modest point of view. As academic researchers we can and must do "no more" than to propose and suggest the concepts, tools and paths that we can derive from our research and reflection.

What should be protected through opacity or privacy tools and what should be protected through transparency tools? What is, in a democratic constitutional society, so essential that it must be as a rule shielded from interference by others (public and private actors)? Which aspects of individual life in an open society must be protected against openness and transparency?[102] Which aspects of individual life should be withdrawn from scrutiny, surveillance and control? Where are hard norms needed? Where should ad hoc balancing be replaced by categorical balancing?

102 This is actually the core question of David Brin's inspiring book D. BRIN, *o.c.*, 378 p. Nonetheless, we defend a different position, inasmuch as we do not carry *mutual transparency* (and *symmetric information flows*) as far as Brin does. We do not value anonymity and opacity so negatively as he does. The fundamental reason for this, we think, is that Brin distinguishes freedom ("personal sovereignty") and privacy much more sharply than we do: for him privacy is "a delicacy that free people can pour for themselves as much or as little as they choose [...] Privacy is a wonderful highly desirable *benefit* of freedom" (p. 79). Brin associates freedom with free speech and comes to the conclusion that "there can be few compromises when it comes to the underpinnings of liberty. Without both individual freedom and distributed sovereignty, all our vaunted modern privacy would vanish into legend" (p. 79). Our understanding of privacy is precisely interwoven with the "underpinnings of liberty", and that is why we tend to give privacy a more positive and broader connotation. For Brin, privacy only concerns a limited array of aspects which come close to the sanctity of the home: "[...] I won't exchange my liberty or anyone else's – for security. I certainly won't give up essential privacy: of home, hearth, and the intimacy that one shares with just a few".

The answers to these questions must be formulated by reference to the basic features of the democratic constitutional state. From this perspective opacity/privacy rules – prohibitionary rules – should guarantee those aspects of an individual's life that embody the conditions for his/her autonomy (or self-determination, or freedom, or "personal sovereignty"). This is the case because it is precisely this autonomy that develops and fuels both one's participation in the civil and political life and the fact that one develops a personality and a social/relational life. Privacy must protect what lies behind the persona, the mask that makes an individual a legal person (cf. anonymity). It must preserve the roots of individual autonomy against outside steering, against disproportionate power balances, precisely because such interference and unbalanced power relations do more than threaten individual freedom; they also threaten the very nature of our societies. Privacy and opacity are needed because, as we have already explained, a democratic constitutional state is primarily concerned with the protection of the individuals' autonomy (and resistance) in vertical, but also in horizontal power relations.

6.3 An example: camera surveillance

Both in Europe and in the United States there is evidence of a common understanding of the impropriety of filming people in private houses. In the early seventies, abuses with spy devices gave way to specific criminal laws on visual intrusion in Italy, the Netherlands and France[103]. Filming what happens inside houses was targeted by very simple, clear-cut prohibitions. On this solid basis, a data protection framework was build to organise protection for people under surveillance in the public space. In 1995, France enacted a much broader bill initiating a system of CCTV-rights and duties, whereby use of CCTV without prior notification to a control board is a criminal deed, as is refusing access to the images and refusing rights of correction and rectification[104]. Due to data protection, countries with older specific criminal bills, such as France, are now able to redefine their scope of protection. In some German Länder specific video-clauses are introduced in local data protection bills. The Schleswig-Holstein bill allows

103 Cf. section 4 of the Italian act respecting workers' freedom of 20 may 1970 (amended in 1990); the French law of 17 July 1970 (article 368 (old) penal code/article 226 new penal code); the Dutch law of 7 April 1971 (art. 139f-g, 372-375 and 441a penal code).

104 Bill n. 95-73 "d'orientation et de programmation relative à la sécurité", 21 January 1995; P. DE HERT, "La police française en l'an 2000", *Politeia. Revue professionnelle Belge des services de police*, 1995, n. 4, pp. 12-16.

EVS of public places when this is necessary for the protection of the premises and when other legitimate interests of a higher rank are not present. In the case of recordings of visual data, there has to be a warning to alert the people concerned[105].

Belgium has a general bill on data protection, but did not and does not have a bill on digital video surveillance or video surveillance. In 1994 the Minister of Justice refused to enact specific legislation with simple prohibitions on certain uses of CCTV because he thought this would inhibit the legitimate business of CCTV, deemed necessary for safety purposes[106]. When, subsequently, the Belgian data protection authority drafted recommendations for the use of CCTV based on the 1992 bill on data protection, it reinvented or imported the French or Dutch "very simple" prohibitions on filming private spaces, without however referring to these prohibitions[107].

This example shows that data protection bills and criminal or other prohibitions do not contradict each other, but have a complementary nature. Different domains of law have different kinds of logic that can be combined, in this case criminal law (prohibitive logic) and data protection (channelling logic). Where law is an appropriate and effective instrument, there is a need to identify the relevant harm with precision in order to craft a precise and targeted solution, capable of carrying basic messages that bring "ontological security" (cf. above). We thus favour a framework that weighs the benefits of different kinds of legal logics against the possible threats to privacy. Using this balancing approach, we advocate narrowly targeted legislation aimed at enhancing the protection of sensitive domains of human life (for instance, processing of medical and financial information). The core of privacy should be clearly defined in terms of harmful uses. Normal processing of data and potential harms must be addressed by data protection with its channelling/procedural logic[108].

Depending on the circumstances, the existing equilibrium between the use of tools of opacity and the use of tools of transparency has to be

105 Cf. section 32 of the Schleswig-Holstein bill "zum Schutz personenbezogener informationen", 30 October 1991.

106 See in detail P. DE HERT, O. DE SCHUTTER & S. GUTWIRTH, "Pour une réglementation de la vidéosurveillance" [Plea for a legal framework for CCTV], *Journal des tribunaux*, 21 September 1996, pp. 569-579.

107 *Ibid.*

108 L. BERGKAMP, "The Privacy Fallacy: Adverse Effects of Europe's Data Protection Policy in an Information-Driven Economy", *Computer Law and Security Report*, 2002, v. 18, n. 1, p. 46.

altered. A fine example is provided by a recent bill in Holland on video monitoring of employees, which prohibits the secret use of a camera to detect fraud, except by the police. This prohibition, which supplements existing prohibitions on filming of private premises and existing rules of data protection governing other use of CCTV, effectively forbids the installation of a hidden camera by a shop-keeper who suspects that an employee is stealing. The new prohibition contains one clear message: from now on, the shopkeeper will have to ask the police to mount such surveillance[109].

6.4 A second example: passenger profiling

In a recent Dutch publication we have carried out a similar exercise with regard to the Computer Assisted Passenger Pre-Screening System (CAPPS II)[110]. We recall that the core idea of this U.S. project is to focus security resources on suspicious travellers, while ensuring that most people are not inconvenienced by this heightened security. The aim of the American initiative is to screen risks before boarding. CAPPS-II is designed to reduce the number of people for further screening and is intended to "restore public trust" in airline travel. It proposes to use extensive data mining of credit history, criminal records, travel patterns and to expand the range of databases searched for suspicious activities to profile *all* airline passengers. The new system uses an algorithm to determine indicators of characteristics or behaviour patterns that are related to the occurrence of certain behaviour. Risk scores then help to determine whether a passenger can board a flight. CAPPS-II allows airport authorities to discover through a computer search whether a person is to be identified as a possible suspect. The information is forwarded to the appropriate law enforcement agencies. Before boarding the plane, passengers are assigned one of three rankings printed in code on their boarding passes: green requires routine security, yellow, "added checks", while red bars passengers from flying and subjects them to law enforcement

109 THIERRY D. M., "Het gebruik van camera's ter opsporing van strafbare feiten op de werkvloer", *Bb* 24 juni 2001, n. 12, pp. 132-134.

110 P. DE HERT & S. GUTWIRTH, "Veiligheid en grondrechten. Het belang van een evenwichtige privacypolitiek" [Security and Human Rights. The Importance of a Balanced Privacy Policy], in E. R. MULLER (red.), *Veiligheid. Studies over inhoud, organisatie en maatregelen*, Alphen aan den Rijn, Kluwer, 2004, 585-630. See also M. V. PÉREZ ASINARI & Y. POULLET, "The Airline Passenger Data Disclosure Case and the EU-US Debate", *Computer Law & Security Report*, 2004, v. 20, n. 2; 98-116; M. V. PÉREZ ASINARI & Y. POULLET, "Airline Passengers" Data: Adoption of an Adequacy Decision by the European Commission. How will the Story End?", *Computer Law & Security Report*, 2004, v. 20, n. 5, 370-376.

investigation. The red code would be reserved for those on terrorist watch lists.

The CAPPS-II initiative has consequences for Europe. At least since 5 march 2003, United States authorities have had access to most European airlines" passenger databases. An agreement between the European Commission and United States Customs gives the USA online access to the passenger name record (PNR) data of all Europe-based airline carriers for flights that go to, from or through the USA. The scope of the agreement is wide. The agreement says that "Customs will retain the data no longer than is required for the purpose for which it was stored". But at the same time it is clear that the data is stored for an almost unlimited number of purposes, certainly not limited to fighting terrorism: "PNR data is used by Customs strictly for enforcement purposes, including use in threat analysis to identify and interdict potential terrorists and other threats to national and public security". The U.S. Customs will also share the data with all other U.S. agencies: "Other law enforcement entities may specifically request PNR information from Customs and Customs, in its discretion, may provide such information for national security or in furtherance of other legitimate law enforcement purposes"[111]. The agreement reads as an assurance that EU passenger data will be stored in FBI, NSA and CIA databases.

The agreement between the European Commission and U.S. authorities on the transmission of PNR data has encountered fierce opposition during a public hearing at the European parliament. During a public hearing on 25 March 2003in the European parliament the Commission argued that it had no choice but to accept the U.S. demands for passenger data. Threats to fine European airlines or even halt landing rights were taken very seriously by the Commission. But many participants were not satisfied with the explanation that the Commission had been blackmailed and couldn't do anything about it. They argued that the transfer of PNR data has no legal basis and is a direct violation of the EU Data Protection Directive. The European Parliament decided, in a unique move, to bring the case before the European Court of Justice. A judgement is expected in 2005.

Our contribution to the debate has been to identify some structural shortcomings in the European responses. Fundamentally, events have

111 European Commission/US Customs talk on Passenger Name Record (PNR) transmission, *Joint Statement of 17/18 February 2003*, see http://europa.eu.int/comm/external_relations/us/intro/pnrjoint03_1702.htm.

proven the limited reach of data protection. Europe applied transparency tools where opacity-like answers were needed. Data protection authorities drew attention very early to the problematic nature of the initiative, but were unable to bring the case to a satisfactory end. The fundamental question was circumvented, namely: do we want these kind of data-mining practices in Europe with all the risks for citizens who match a profile? Politicians, not surprisingly, had to take over and phrase the debate in these elementary terms. Unable to per suade the European Commission to adopt a firm position, they finally referred the matter to the European Court of Justice. We can only hope that this Court will effectively deal with the crucial issue, viz. the desirability or necessity of these projects for modern democratic states. It might very well be that the outcome will be a positive "go" and then data protection can do its proper job and clean up the more procedural problems with the American project.

6.5 Workable criteria?

It is possible for us to be more specific about the use of the respective tools or do we abdicate? When are opacity tools necessary? Only a beginning of the answer is possible.

Firstly, we think that opacity tools might be needed with respect to the set of values protected through the inviolability or sanctity of the home. People need places were they can rest and come to terms with themselves in a sphere of trust and security, in an environment where they experience "ontological security" – a place where family life is possible. Such places represent a private territory, a sanctuary; they imply intimacy, anonymity and a possibility of solitude. The protection of the home is one of the oldest human rights and nowadays it is still enshrined in international human rights law and in national constitutions. For these legislations home is where the house is, or by extension, the car, the caravan, [...] Their terms refer to the material/physical home. But, indeed, "being at home" means more than being in a certain physical environment, it means also feeling at ease, comfortable. In French the concept of "home" can also be translated by *chez soi* which is actually expressive from our point of view. It is the *chez soi* that a democratic constitutional state respects and values (contrary to totalitarian states): the idea that one can be with him/herself without outside interference in order to "manage" his/her relational, civil and political life as a free being. In a certain sense it is precisely this *chez soi* that lies protected behind the mask of legal personality. Values protected by the inviolability

of the home therefore might thus also need protection outside the material house, and especially in the "virtual world".

Secondly, tools of a prohibitive nature are obviously required when other firmly rooted (in tradition or in law) human rights are at stake, such as the right to have correspondence and the content of communication protected (cf. *supra*). Disregard of these rights has brought legislation in the United Kingdom giving the employer almost absolute discretion to monitor destinations and control of e-mail sent by employees[112]. This stands in complete contrast with the approach followed in Belgium, where concerns about the right to have correspondence respected has inspired a regulation whereby the employer can check on the destination and on other data about the telecommunication, but cannot check on the content of the communication[113].

Thirdly, opacity exist within the framework of data protection, *e.g.* in the case of sensible data or in the case of decisions regarding individuals taken solely on the basis of automatic profiling. Indeed, the additional danger here is linked to possible discriminatory effects of such practices.

Fourthly, a need for opacity can be drawn from the function of human rights in promoting and encouraging citizenship. When certain law enforcement tactics threaten aspects of human behaviour vital to the formation of the free and equal citizen, a prohibitive logic will impose itself.

Finally, opacity tools can be implemented when there is a societal interest in ending ongoing debates and a need to enhance legal certainty.

Having tentatively set out some touchstones which we deem to be relevant for identifying issues that demand a privacy/opacity legal approach, it is easier to point out what, at the other side of the switch, will fall under the data protection/transparency legal approach. We are tempted to answer: in principle, all forms of processing of personal data that do not demand categorical balancing by the (constitutional) legislator on the basis of the guidelines identified above. In reality this residual category may well have a considerable dimension, especially in horizontal relations where the

112 K. BEST & R. MCCUSKER, "The Scrutiny of the Electronic Communications of Businesses: Striking the Balance Between the Power to Intercept and the Right to Privacy?", *Web Journal of Current Legal Issues*, 2002/1, via http://webjcli.ncl.ac.uk/2002/issue1/kb-rm1.html, pp. 9-11.

113 P. DE HERT, "C.A.O. n. 81 en advies n. 10/2000 over controle van Internet en e-mail" [Labour law: Soft law on e-mail and Internet practices], *Rechtskundig weekblad,* 2002-2003, v. 66/33 (19 April 2003), pp. 1281-1294.

existence of unequal power relations should not be taken as a starting point. In all cases where consent (still) plays an important role, it can be assumed that the guidelines for the opacity approach are not met. Individual responsibility, consent and ad hoc balancing are sufficient but indispensable conditions for meeting the requirements of a constitutional state. Correlatively, this implies that secret processing of personal data must be prohibited, since this secret or unknown character renders the mise en oeuvre of transparency tools or data protection impossible. Informed consent is a *conditio sine qua non* for subsequent controls on data processing[114]. There is little room for a policy based on the full liberty of processing data. Transparency tools are the default tools in all areas of data processing. Because this processing of data is always intertwined with power relationships[115], a democratic constitutional state must foresee rules that permit to channel this power, namely rules of transparency and accountability, viz. data protection rules.

CONCLUSION

The development of the democratic constitutional state has led to the invention and elaboration of two complementary sorts of legal tools, namely normative opacity tools that draw the limits of interference with individuals, and transparency tools that organise the channelling, control and restraint of power. Privacy is an example of the former, data protection and criminal procedure are mainly examples of the latter. Hence, privacy is about much more than just accountability and foreseeability. This is the reason why case-law and literature, in our opinion, tend to overstress the importance of accountability and foreseeability relating to privacy limitations to the detriment of the normative and prohibitive drawing of barriers. There is too much "yes, if" and a lack of "no". By recognising human rights, the democratic constitutional state generated legal mechanisms to impose non-interference in the private sphere. In our opinion, privacy is closely linked to this endeavour as it tends to protect the values of liberty by erecting a legal shield against interferences. More concretely this shield can take the form of (legal) claims of immunity, anonymity, pseudonymity, opacity and sanctity.

114 Cf. The fairness principle of the OECD guidelines and Treaty 108.

115 Why? Because the processing of information about individuals always opens the door to some form of control. Because it intimidates. Because people adapt their behaviour if it is clear that information is gathered. This is not intrinsically a "bad thing", but just a fact that must be taken seriously into account.

This normative position is echoed in the legal landscape. The right to privacy is a widely recognised opacity tool to prohibit certain uses of power. It may not be the strongest human right listed in the ECHR and it may well also be that the "reign" of privacy in discourse is over, but nevertheless the right is there and it has its proper place, a quintessential place. The interpretative work of the European Court on Human Rights is ambiguous. Obviously, the European judges prefer to concentrate on non-privacy issues such as accountability and foreseeability of privacy limitations. Gradually the Court has developed the view that a legal basis for privacy infringements should not only exist, but should also meet some qualitative requirements, namely accessibility and foreseeability. Also the Court has fused in its analysis elements borrowed from procedural rights enshrined in other provisions of the Convention. But the question whether a certain practice is "necessary in a democratic society" is often not addressed by the Court when a breach of the legality requirement is found. Constitutional reasonableness encompasses substantive fairness *and* procedural regularity and the two are often tightly intertwined, but in a sensible division of constitutional labour these issues are best treated separately on their own merits. Searches with a maximum of procedural guarantees might well be unreasonable in the light of Article 8, when they allow for an unwanted concentration of power (the idea behind opacity tools). On the other hand it is clearly possible to conceive privacy infringements without procedural guarantees that are nevertheless reasonable, for instance metal detectors at airports[116]. Elaborating codes of criminal procedure, for instance, with regard to telephone tapping, is not a command that follows from Article 8.

From that perspective the constitutionalisation of the right to data protection in the EU Charter represents a positive evolution, for data protection has a precise target which is distinct from privacy concerns and aims at organising the fair processing of personal data by both public and private actors. That is why the precise requirements imposed upon the processing of personal data have been developed by data protection law and not by the constitutional construction of privacy. Moreover, due to its channelling logic data protection allows for a sensible constitutional division of labour. Its recognition may pave the way for a rediscovery of the power blocking nature of the privacy right. Privacy and data protection

116 A. R. AMAR, *The Constitution and Criminal Procedure. First Principles*, New Haven and London, Yale U.P., 1997, pp. 38-39.

supplement each other and pre-suppose each other. We have suggested that the European legislator considers both tools and uses them accordingly. By giving concrete examples (*e.g.* video surveillance and CAPPS-II) and by identifying guidelines for switching from the privacy logic to the data protection logic and back, this contribution aims at illuminating how Europe is *and* should be dealing with challenges created by new power demands, especially in the field of law enforcement.

Privacy functions as a tool to distinguish between the reasonable and the unreasonable and to stop law enforcement power whenever it crosses a normative line. Privacy is not a tool to regulate reasonable law enforcement power by devising mechanisms of transparency, control and accountability. Opacity and transparency each have their own role to play. They are not communicating vessels. Hence, unlike Etzioni, we do not think that public authorities cannot be denied technologies and means for crime fighting if their implementation is linked to enough transparency and accountability[117]. On the contrary, privacy implies the making normative choices: some intrusions are just too threatening for the fundamentals of the democratic constitutional state to be accepted even under a stringent regime of accountability. Other intrusions, however, will be felt to be acceptable and necessary in the light of other sometimes predominating interests. Only then, after such a normative weighing of privacy and other interests, privacy invasive and liberty threatening measures can be, exceptionally and regrettably, accepted and submitted to the legal conditions of transparency and accountability.

117 A. ETZIONI, "Implications of Select New Technologies for Individual Rights and Public Safety", *Harvard Journal of Law & Technology*, 2002, v. 15, n. 2, p. 34: "If accountability is deficient, the remedy is to adjust accountability, not to deny the measure altogether".

Sobre os autores:

Serge Gutwirth | *E-mail:* serge.gutwirth@vub.ac.be

Professor at the Faculty of Law and Criminology of the VUB, where he studied law, criminology and also obtained a post-graduate degree in technology and science studies. Today he is mainly focussing on the difficult articulations between law, politics, technology and ethics with regards to concrete emerging issues (*e.g.* data protection, S&M, gene editing, [...]). Since 2017 he started doing research on the resurgence of the commons. In 1999 Gutwirth has founded the VUB-Research group on human rights (HUMR), which he chaired until 2003, after which his colleague Paul De Hert took over. After obtaining the fellowship in October 2003 he founded the VUB-Research group Law Science Technology & Society (LSTS) which still (co-)chairs. Today, with more than 30 researchers at all levels of experience, LSTS has become a prominent European research institute. Next to thisGutwirth has been Vice-Dean of the Faculty (2012-2018), Vice Chair of the VUB's Research Council from 2002 to 2013, (co-)Director of LSTS (2003-[...]) and Chair of the Department of Interdisciplinary Studies of the law (JURI) from October 2018 on.

Paul De Hert | *E-mail:* pdehert@law.leidenuniv.nl

Prof. Paul De Hert's work addresses problems in the area of privacy & technology, human rights and criminal law. A human rights approach combined with a concern for theory is the common denominator of all his work. In his formative years, De Hert studied law, philosophy and religious sciences (1985-1992). He is Director of the Research group on human rights (FRC) and Vice-Dean of the Faculty and former Director of the Research group Law Science Technology & Society (LSTS), and of the Department of Interdisciplinary Studies of Law.He is board member of several Belgian, Dutch and (other) international scientific journals such as The Computer Law & Security Review (Elsevier), The Inter-American and European Human Rights Journal (Intersentia) and Criminal Law & Philosophy (Springer). He is co-editor in chief of the Supranational Criminal Law Series (Intersentia) and the New Journal of European Criminal law (Sage). Since 2008 he has edited with Serge Gutwirth, Ronald Leenes and others annual books on data protection law (before Springer, now Hart) that, – judging sales numbers, quotations and downloads, attack a massive readership and have contributed to creating the legal, academic discipline of data protection law. De Hert is now series editor of The Computers, Privacy and Data Protection series, now published by Hart.

Artigo convidado.

Cuidar Ou Delatar? A Violação do Sigilo do Prontuário Médico na Criminalização de Mulheres por Aborto Autoprovocado no Estado do Paraná (2017 a 2019)

To Take Care Or to Report? Violation of the Confidentiality of the Medical Record in the Criminalization of Women for Self-Induced Abortion in the State of Paraná (2017 to 2019)

KATIE SILENE CÁCERES ARGUELLO[1]
Universidade Federal do Paraná (UFPR).

VANESSA FOGAÇA PRATEANO[2]
Universidade Federal do Paraná (UFPR).

RESUMO: O presente artigo tem como objetivo investigar e analisar autos judiciais que tratam da investigação e do processamento de mulheres pelo crime de aborto autoprovocado (art. 124 – primeira parte do Código Penal brasileiro) para identificar a participação de profissionais de saúde e o uso dos dados do prontuário médico para a captura das pacientes pelo Sistema de Justiça Criminal logo após procurarem atendimento hospitalar no pós-abortamento clandestino e inseguro. Utilizou-se o método indutivo por meio de uma pesquisa documental e bibliográfica. Foram analisados 43 autos judiciais em trâmite em 15 comarcas do estado do Paraná entre 2017 e 2019. Os resultados demonstraram que 44% das mulheres foram reportadas à polícia por profissionais de saúde; 65% tiveram seu prontuário médico compartilhado com a autoridade policial, sem o seu consentimento; em 58% dos casos em que a mulher foi denunciada à Justiça, tais profissionais de saúde foram arroladas(os) como testemunhas de acusação; e 84% das mulheres delatadas por profissionais de saúde foram atendidas via Sistema Único de Saúde (SUS). Mesmo que a normativa jurídica nacional e internacional proíba tal conduta, por se tratar de violação do sigilo do prontuário médico e, portanto, de direitos fundamentais dessas pacientes, em apenas 16,5% dos casos a defesa abordou e

1 Orcid: https://orcid.org/0000-0001-9360-293X.
2 Orcid: https://orcid.org/0000-0003-3669-6671.

questionou tal prática, e em nenhum dos casos a temática foi analisada desde uma perspectiva de gênero. Conclui-se que a prática é disseminada e pouco questionada pela defesa das mulheres, o que acaba por gerar uma invisibilização da violação dos direitos das mulheres criminalizadas por aborto no âmbito do processo penal.

PALAVRAS-CHAVE: Aborto; processo penal; gênero; sigilo do prontuário médico; pesquisa empírica em Direito.

ABSTRACT: The present work aims to investigate and analyze court records dealing with the investigation and prosecution of women charged with the crime of self-induced abortion (Article 124 – first part of the Brazilian Penal Code) to identify the participation of health care professionals and the use of data from the medical record for the capture of those patients by the Criminal Justice System soon after seeking for hospital care in post-clandestine and unsafe abortion. The inductive method was used through documentary and bibliographic research. Forty-three court records processed in 15 counties in the state of Paraná between 2017 and 2019 were researched. The results showed that 44% of the women were reported to the police by health professionals; 65% had their medical records delivered to the police authority without their consent; in 58% of the cases where a woman was brought to justice, such health professionals were listed as witnesses for the public prosecution; and 84% of the women reported to the police by health care professionals were assisted by the Sistema Único de Saúde (SUS). Despite the fact that national and international legal regulations prohibit such practice, since it is a violation of the confidentiality of the medical record and, therefore, of the fundamental rights of these patients, in only 16.5% of the cases the defense addressed and questioned such practice, and in none of the cases the topic was analyzed from a gender perspective. We can conclude that such practice is widespread and little questioned by the women's defense, which renders invisible the violation of the rights of women criminalized with the crime of abortion within the scope of criminal proceedings.

KEYWORDS: Abortion; criminal proceedings; gender; medical record confidentiality; empirical research in Law.

1 APONTAMENTOS INICIAIS

O aborto[3] no Brasil é criminalizado à exceção de três situações: quando visa a interromper gravidez resultante de estupro; quando o procedimento é a única forma de salvar a vida da gestante, ou seja, quando a continui-

3 Ressaltamos, desde já, que aqui os termos "aborto" e "abortamento" são utilizados como sinônimos, já que o termo "aborto" é o utilizado pelo Código Penal brasileiro; no entanto, do ponto de vista técnico, "abortamento" é a interrupção da gravidez até a 20ª ou 22ª semana e com produto da concepção pesando menos que 500g, e "aborto" é o produto da concepção eliminado no "abortamento" (Brasil, 2011, p. 29).

dade da gestação implicar sua morte; e nos casos em que o feto carregado pela mulher é comprovadamente anencéfalo, condição que impossibilitará sua vida fora do útero. Fora desses enquadramentos jurídicos, a mulher que realizar aborto em si mesma incorrerá no crime previsto no art. 124, primeira parte, do Código Penal de 1940, e estará sujeita a uma pena de detenção de um a três anos.

Embora o ato de induzir um aborto em si mesma seja crime há 81 anos, é, ao mesmo tempo, uma prática corrente no País, já que pesquisas nacionais demonstram que uma a cada cinco mulheres com a idade de até 40 anos já realizou ao menos um aborto (Diniz; Medeiros; Madeiro, 2016)[4]. No que diz respeito às pesquisas que tratam especificamente dos processos de criminalização por aborto autoprovocado, observa-se que tais mulheres, em sua maioria, são capturadas pelo Sistema de Justiça Criminal (SJC) após enfrentarem complicações de um procedimento que "deu errado", procurarem por socorro em uma unidade de pronto-atendimento hospitalar e serem delatadas à polícia pelas(os) mesmas(os) profissionais de saúde que as atenderam e deveriam lhes prestar cuidados[5].

Tais pesquisas são recentes e ainda em pequeno número, dada a invisibilidade ou a generalização que marcam os processos de criminalização que afetam majoritária ou exclusivamente as mulheres não só no Brasil, mas também no mundo, o que ensejou, inclusive, o nascimento da Criminologia Feminista como um campo que visa a denunciar e erradicar tal problema, reproduzido não apenas pela criminologia tradicional, mas pela própria criminologia crítica (Campos, 2017; Mendes, 2017; Ngaire, 1997; Smart, 1976).

Nos últimos 20 anos, pesquisas realizadas em São Paulo, Rio de Janeiro e Distrito Federal, a título de exemplo, demonstraram que, embora a regra legal e ética determine a incomunicabilidade entre a esfera criminal e a esfera da saúde, no que se refere aos processos de criminalização por aborto induzido, a maioria das mulheres passa a ser investigada por

4 É necessário mencionar que tal pesquisa entrevistou apenas mulheres alfabetizadas (era necessário responder a um questionário) e moradoras de zonas urbanas; logo, as taxas podem ser ainda maiores.

5 Mesmo quando as mulheres não são delatadas pelas(os) profissionais de saúde, ainda sofrem estigmatização e são maltratadas quando as equipes notam ou desconfiam que se trata da prática de um aborto fora das hipóteses legais. Mesmo nos casos em que não há confissão da mulher, e ela alega que o aborto foi espontâneo ou resultado de acidente, prevalece um regime de suspeição sobre a sua palavra. Uma pesquisa da Fundação Perseu Abramo (2010, p. 191) demonstrou que 53% das mulheres entrevistadas disseram ter sofrido alguma forma de violência durante a assistência médica após o aborto; 34% foram questionadas insistentemente se haviam induzido o aborto e tratadas como suspeitas de tal crime; 17% das entrevistadas foram acusadas de terem cometido crime e ameaçadas de serem denunciadas à polícia; 16% passaram horas à espera da informação sobre se seriam internadas; e 5% das mulheres foram expostas ao feto e interpeladas com palavras como "olha o que você fez!".

tal conduta a partir de uma comunicação feita à polícia por médicas(os), enfermeiras(os), psicólogas(os) e assistentes sociais que atuam em unidades de pronto-atendimento de hospitais onde tais mulheres dão entrada com hemorragias, fortes dores e lacerações nos órgãos (Ardaillon, 2000; Gonçalves; Lapa, 2008; Cunha; Noronha; Vestena, 2012; Rio de Janeiro, 2018; Melo, 2018; São Paulo, 2018; Ribeiro, 2020).

O cenário vislumbrado por nossa pesquisa não se mostrou diferente daquele encontrado nos trabalhos acima citados; buscamos analisar o total de autos judiciais de mulheres investigadas e processadas criminalmente por aborto autoprovocado em tramitação no estado do Paraná entre 1º de janeiro de 2017 a 31 de dezembro de 2019, e chegamos a um *corpus* empírico de 43 autos em 15 comarcas. A constatação de que quase metade das mulheres foi capturada pelo SJC após comunicação do suposto crime feita por profissionais de saúde às autoridades policiais e de que mais da metade teve as informações de seu prontuário médico – que são sigilosas – compartilhadas com a polícia sem o seu consentimento suscitaram discussões éticas, criminais, processuais penais e constitucionais, e também de gênero, a respeito de tal prática.

No presente artigo, portanto, buscamos analisar criticamente e sob a perspectiva de gênero o início de investigações policiais e o processamento penal de mulheres por aborto autoprovocado a partir de comunicação feita por profissionais de saúde às autoridades policiais, assim como o uso das informações do prontuário médico dessas mulheres para fundamentar o seu indiciamento ou denúncia, requisitadas pela polícia ou pelo juízo sem autorização da paciente e muitas vezes fornecidas pela instituição e pelas(os) profissionais de saúde de forma voluntária, como se pode observar em nossa pesquisa.

Do ponto de vista metodológico, o artigo está dividido em três partes. Na primeira parte, apresentamos o quadro normativo brasileiro a respeito do sigilo profissional e do fenômeno do uso dos dados do prontuário médico para fins criminais, por meio de pesquisa bibliográfica junto à normativa constitucional, penal, processual penal e também junto ao Código de Ética Médica, a fim de se problematizar as práticas policiais e judiciais de uso de tais informações sem o consentimento de sua titular, no caso, a paciente criminalizada por aborto autoprovocado.

Em um segundo momento, apresentamos resultados de pesquisa documental (Reginato, 2017; Silva, 2017) de abordagem qualitativa (Yin, 2016) e quantitativa realizada junto aos supracitados 43 autos judiciais, em

que são apresentados a forma como se deu a captura de tais mulheres pelo SJC e também o itinerário[6] dessa criminalização.

Posteriormente, em uma terceira fase, a partir do aporte da Criminologia Feminista (Campos, 2017; Mendes, 2017; Ngaire, 1997; Smart, 1976) e de pesquisas no âmbito da saúde coletiva, buscar-se-á problematizar de que forma o uso dessas informações para a criminalização de mulheres compromete o exercício de vários direitos fundamentais por parte da população feminina brasileira, como o direito à intimidade, à saúde, à não discriminação no acesso à saúde e à não autoincriminação.

2 O ENQUADRAMENTO DO SIGILO MÉDICO NA NORMATIVA JURÍDICA BRASILEIRA

Para se identificar o percurso de construção da categoria do sigilo médico na normativa jurídica brasileira, deve-se iniciar pela leitura do próprio texto constitucional.

A Constituição Federal de 1988, em seu art. 5º, X, dispõe sobre a inviolabilidade da intimidade, da vida privada, da honra e da imagem, conferindo a tal vedação o *status* de uma garantia fundamental da pessoa humana. Em seguida, ainda no art. 5º, notamos o inciso XIV, que garante a todas(os) o acesso à informação, porém resguarda o sigilo da fonte quando necessário ao exercício profissional. Por fim, no mesmo artigo, inciso LXIII, há a consagração do princípio de que ninguém pode ser compelido a produzir prova contra si mesma(o).

É possível notar o diálogo entre tais garantias fundamentais no sentido de que alguns profissionais, intituladas(os) "confidentes necessárias(os)", notadamente as(os) profissionais de saúde, necessitam de acesso à esfera mais íntima de cada indivíduo para que possam exercer, de forma efetiva, a sua função; de outro lado, aquela(e) que garante acesso à sua esfera mais

6 Adota-se aqui a expressão "itinerário da criminalização" por meio de um diálogo deste trabalho com estudo realizado pela antropóloga Débora Diniz e pelo sociólogo Marcelo Medeiros, que, em sua Pesquisa Nacional do Aborto (2010 e 2016), buscam descortinar o itinerário do aborto no País. Nossa pesquisa é mais afunilada, no sentido de que não nos propomos a identificar e analisar o cenário da prática do aborto no estado do Paraná, mas o itinerário da sua *criminalização*, partindo-se do pressuposto já demonstrado pela Criminologia Crítica de que há um enorme fosso a separar a prática do que é considerado crime do processo de criminalização, que é seletivo e neutralizante, deixando para trás a maioria daqueles e daquelas que praticaram atos considerados criminosos, fatos que fazem parte da chamada "cifra oculta do crime". Como diria Winfried Hassemer, a diferença entre o criminoso e o não criminoso é que o primeiro foi descoberto. "Pode-se dizer de modo acentuado que as teorias do autor não descrevem os *criminosos*, mas os *descobertos*, e que os outros criminosos, em todo caso, se distinguem dos descobertos por um ponto altamente significativo: eles conseguiram permanecer ocultos no setor obscuro" (Hassemer, 2005, p. 101).

íntima o faz por uma necessidade, em geral para o exercício de direitos fundamentais, como o direito à saúde ou à ampla defesa. Tal acesso, no entanto, deve estar protegido de invasões por parte de terceiros, da sociedade e até mesmo do Estado, motivo pelo qual as informações compartilhadas neste contexto estão protegidas pelo sigilo.

No direito privado, o direito à intimidade e à privacidade, por envolver um importante aspecto dos direitos da personalidade, encontra expressão nos arts. 388, II, e 448, II do Código de Processo Civil (a partir da revogação do art. 229 do Código Civil de 2002), os quais determinam, respectivamente, que a parte e a testemunha não são obrigadas a depor sobre fatos a cujo respeito, por estado ou profissão, devam guardar sigilo. Ainda, o CPC escusa do dever de exibir documentos ou coisas em juízo a(o) profissional (seja como terceiro, seja como parte) que deve guardar segredo, de acordo com o art. 404, IV.

Na seara penal, o Código Penal de 1940 tipifica, em seu art. 154, o crime de violação de segredo profissional, que consiste em revelar, sem justa causa, segredo de que se tem ciência em razão de função, ministério, ofício ou profissão, e cuja revelação possa produzir dano a outrem, sendo um ato punível com pena de detenção de três meses a um ano, ou multa. Vê-se, portanto, que o ato é considerado crime, embora de menor potencial ofensivo.

A Lei das Contravenções Penais de 1941, em seu art. 66, II, traz a contravenção que diz respeito a deixar de comunicar à autoridade competente um crime de ação pública, do qual a pessoa teve conhecimento no exercício da Medicina ou de outra profissão sanitária, *desde que a ação penal não dependa de representação e a comunicação não exponha o cliente a procedimento criminal*. Ou seja, a comunicação ao SJC, caso venha a expor o(a) cliente (ou paciente) à criminalização – mesmo que o crime em tese cometido seja processado mediante ação pública incondicionada –, não deve ser realizada.

Já o Código de Processo Penal, em seu art. 207, estabelece que "são proibidas de depor as pessoas que, em razão de função, ministério, ofício ou profissão, devam guardar segredo, salvo se, desobrigadas pela parte interessada, quiserem dar o seu testemunho" (Brasil, 1941). Ou seja, há a vedação do depoimento de tais profissionais, seja em sede policial ou já na fase processual, salvo se a(o) titular das informações conceder autorização para tanto.

A Lei Geral de Proteção de Dados (LGPD), que passou a vigorar em 2020, em seu art. 5º, II, trata os dados referentes à saúde de pacientes como dados pessoais sensíveis, e determina, em seu art. 7º, VIII, que o tratamento de tais dados só pode ocorrer exclusivamente no sentido de tutelar a saúde, o que nos permite concluir que não se admite o tratamento deles para fins de persecução penal da(o) paciente.

No que diz respeito aos códigos deontológicos das profissões de saúde, a questão do sigilo ou segredo profissional é tratada de forma expressa, de forma mais ou menos aprofundada a depender da normativa, mas sempre sob a determinação de que o compartilhamento de informações da(o) paciente sem a sua autorização constitui infração ética – de acordo com o Código de Ética Médica, por exemplo, em seu art. 73, parágrafo único, a proibição permanece mesmo quando o fato é público, a(o) paciente já é falecida(o), a(o) profissional é intimado a depor como testemunha (o Código afirma que, nessa hipótese, a(o) médica(o) comparecerá perante a autoridade e declarará seu impedimento) ou há suspeita de que um crime foi cometido pela(o) paciente.

Portanto, observa-se que a normativa jurídica brasileira entende como direito fundamental do indivíduo não ter os seus dados médicos compartilhados com terceiros sem a sua expressa autorização, ao garantir o sigilo no exercício de certas profissões, como a de saúde; tal direito mantém-se mesmo diante da persecução penal estatal, uma vez que profissionais de saúde são proibidas(os) de depor sobre fatos de que tomaram conhecimento no exercício de sua profissão. Tal vedação aponta que nem mesmo mediante determinação judicial pode a(o) profissional de saúde violar o sigilo, limitando-se ele(a) a comparecer em juízo e declarar o seu impedimento.

Logo, a conclusão é de que a requisição de delegada(o) de polícia ou determinação judicial para que o prontuário seja entregue, ou para que a(o) profissional de saúde declare o que sabe a respeito dos fatos, não integra o conteúdo de "motivo justo" ou "justa causa"[7], exceções comumente acionadas para se impor o compartilhamento das informações. Tampouco se inclui na exceção de "dever legal", que envolve apenas as situações, previstas

7 O Código de Ética Médica é explícito ao afirmar que o sigilo não deve ser violado nem mesmo em caso de indício ou confirmação de que um crime foi cometido. Há motivo justo ou justa causa quando o médico toma conhecimento de que sua(seu) paciente pretende cometer um fato delituoso, ou seja, abrange casos que possam ocorrer no futuro, mas exclui fatos ocorridos no passado.

em lei, nas quais tal profissional é obrigada(o) a comunicar as autoridades externas às autoridades sanitárias a respeito de suspeita ou confirmação de atos delituosos[8].

No que diz respeito às consequências de tal conduta vedada, está-se diante da produção não apenas de uma prova ilegítima (em violação à norma processual penal anteriormente descrita), mas também ilícita (em decorrência de afronta à garantia constitucional de respeito à intimidade e à privacidade, ao art. 154 do Código Penal e às regras deontológicas profissionais[9], normas de direito material). Assim, em obediência ao art. 5º, LVI, da Constituição Federal[10] e ao art. 157 do Código de Processo Penal[11], tais provas devem ser desentranhadas dos autos, não sendo possível a sua reprodução, por restarem comprometidas e serem passíveis de contaminar todo o processo.

3 A ESFERA CRIMINAL E A ESFERA DA SAÚDE: O QUE OS AUTOS NOS DIZEM SOBRE A RELAÇÃO ENTRE AMBAS?

Para investigar e compreender a participação de profissionais de saúde e o uso dos dados do prontuário médico na captura das pacientes pelo Sistema de Justiça Criminal logo após procurarem atendimento no pós-abortamento clandestino e inseguro no estado do Paraná, foram analisados autos judiciais em tramitação em um total de 15 comarcas paranaenses durante um marco temporal de três anos – de 1º de janeiro de 2017 a 31 de dezembro de 2019. Tratou-se de pesquisa empírica de natureza documental qualitativa e quantitativa.

8 O dever legal de que falam os códigos deontológicos das profissões de saúde está adstrito a casos em que a comunicação ao Sistema de Justiça Criminal está prevista em lei, como situações em que há suspeita ou confirmação de violência praticada contra criança, adolescente, idoso ou pessoa com deficiência, conforme o Estatuto da Criança e do Adolescente (art. 13 da Lei nº 8.069, de 1990), o Estatuto do Idoso (art. 19 da Lei nº 10.741, de 2003) e o Estatuto da Pessoa com Deficiência (art. 26 da Lei nº 13.146, de 2015).

9 O conteúdo dos códigos deontológicos impõe um dever à respectiva categoria profissional que se propõe a regular, ou seja, trata-se de um dever porque os códigos de ética são lei em sentido lato; cada código, embora não seja uma lei no sentido estrito da palavra (votada e aprovada pelo Poder Legislativo), é uma lei jurídica com a denominação técnica de resolução, oponível a todos os profissionais daquela categoria. É, portanto, norma imperativa, oponível a toda a classe profissional e com força coercitiva e de sanção. O Código de Ética Médica, a título de exemplo, está previsto no art. 30 da Lei nº 3.268, de 1957. Não se trata, portanto, de mero ato administrativo; tem natureza de lei (Oliveira, 2001, p. 143).

10 Art. 5º: "inciso LVI – são inadmissíveis, no processo, as provas obtidas por meios ilícitos" (Brasil, 1988).

11 "Art. 157. – São inadmissíveis, devendo ser desentranhadas do processo, as provas ilícitas, assim entendidas as obtidas em violação a normas constitucionais ou legais." (Brasil, 1941)

Nossa hipótese era de que o cenário do estado paranaense não apresentaria grandes distinções em relação àqueles observados em outras pesquisas a respeito do tema, no que diz respeito à contribuição expressiva de profissionais de saúde para a captura dessas mulheres pelas teias do SJC e ao compartilhamento, pelos estabelecimentos de saúde, dos dados do prontuário médico dessas pacientes com as autoridades policiais, sem que elas tenham dado seu consentimento.

Assim, a partir da análise dos 43 casos, foi possível identificar que 19 mulheres foram criminalizadas a partir de comunicação feita à Polícia Civil ou à Polícia Militar por médicas(os), assistentes sociais e enfermeiras(os), o que corresponde a 44,1% do total – cenário não muito distante de pesquisas anteriores realizadas em outros estados da federação e no Distrito Federal –, conforme tabela 1.

Tabela 1: A origem da criminalização de mulheres por aborto no Estado do Paraná (2017 a 2019):

Origem da criminalização	*Casos*	*Proporção*
Denúncia de profissional de saúde	19	44,18%
Denúncia do ex-parceiro	7	16,27%
Denúncia anônima	5	11,61%
Denúncia de terceiros	3	7%
A própria mulher confessou o fato	3	7%
Informação prestada por familiares sobre um aborto espontâneo ou achado de feto na residência	2	4,65%
Informação prestada pela própria mulher sobre um aborto legal ou espontâneo	2	4,65%
Denúncia de familiares	1	2,32%
Suposto fato apareceu em uma investigação policial sobre terceiro	1	2,32%
Total	*43*	*100%*

Fonte: As autoras (2021).

Quando se buscou analisar a modalidade de atendimento hospitalar das 19 mulheres que foram denunciadas ao SJC por profissionais de saúde, observou-se que 16 (ou 84,21% do total) foram atendidas pelo Sistema Único de Saúde (SUS); em um caso, a mulher foi atendida via plano de saúde em um hospital universitário e filantrópico que atende majoritariamente pelo SUS; e, em dois casos, não foi possível identificar se a mulher foi atendida via SUS, plano de saúde ou na modalidade particular, conforme tabela 2.

Tabela 2: A modalidade de atendimento no pós-abortamento de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Modalidade de atendimento da mulher no pós-abortamento	*Casos*	*Proporção*
Sistema Único de Saúde (SUS)	16	84,22%
Plano de saúde	1	5,26%
Não informado	2	10,52%
Total	*19*	*100%*

Fonte: As autoras (2021).

Tanto a forma de entrada da mulher no SJC (majoritariamente via denúncia feita por profissionais de saúde) quanto a modalidade de atendimento prestada a essas mulheres (via SUS) dialogam com achados de outros estudos:

> Em geral, o perfil da mulher se repetia: pobre, pouco instruída, moradora de periferia. Contudo, este não é necessariamente o perfil das mulheres que fazem aborto, mas sim o perfil das mulheres que são presas por terem feito aborto. Deste aspecto percebe-se uma grande diferença. *O sistema captura apenas algumas mulheres, as que necessitam se submeter à saúde pública*. Aquelas que encontram outras soluções, privadas, não são atingidas. Um claro retrato do recorte socioeconômico. (Cunha; Noronha; Vestena, 2012, p. 32 – grifos nossos)

Outro contexto que se buscou analisar diz respeito à obtenção de acesso, pelo(a) delegado(a) de polícia, aos dados do prontuário médico da mulher, ou seja, a um documento com informações que podem configurar indícios ou prova da prática, em tese, do crime de aborto.

Em 30 dos 43 casos estudados (62,8% do total), o prontuário foi requisitado pela Polícia Civil sem apresentação de autorização por escrito da mulher, sendo que, em 28 casos (65,11% do total), o documento foi enviado pelo estabelecimento de saúde e, portanto, efetivamente acessado pela autoridade policial; nos dois casos restantes, respectivamente, o estabelecimento de saúde informou não ter encontrado o prontuário em seus arquivos; e o documento não foi juntado aos autos, não havendo menção a ele. Em sete casos, a mulher concedeu autorização de acesso ou ela mesma compareceu à delegacia com o documento em mãos; nos demais seis casos, o prontuário não foi requisitado, conforme tabela 3.

Tabela 3: O acesso ao prontuário médico de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Acesso ao prontuário	*Casos*	*Proporção*
Sem autorização da paciente	28	65,11%
Com autorização da paciente	7	16,28%
Houve pedido de acesso sem autorização da paciente, mas o prontuário não foi enviado	2	4,65%
Não houve pedido de acesso	6	13,96%
Total	*43*	*100%*

Fonte: As autoras (2021).

Em relação às sete mulheres que concederam o acesso por escrito ou compareceram à delegacia para entregar a documentação, convém mencionar que seis não tiveram acesso à defesa no momento em que concederam tal autorização, já que não estavam acompanhadas de advogado(a) ou defensor(a) público(a), seja no momento do interrogatório, seja no momento anterior ao seu comparecimento espontâneo à delegacia, quando elas ou seus familiares foram orientados pelo estabelecimento de saúde a comparecer à unidade policial para comunicar um aborto espontâneo, conforme tabela 4.

Tabela 4: Informações sobre o acesso à defesa em sede policial e autorização de acesso ao prontuário de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Mulher autorizou acesso ao seu prontuário	*Casos*	*Proporção*
Teve acesso à defesa em sede policial	1	14,28%
Não teve acesso à defesa em sede policial	6	85,72%
Total	*7*	*100%*

Fonte: As autoras (2021).

Por fim, objetivou-se identificar a presença de profissionais de saúde como testemunhas do possível crime de aborto, seja em sede policial, seja em juízo.

Assim, observou-se que, dos 43 casos pesquisados, em 16 (37,2% do total) deles, ao menos um(a) profissional de saúde compareceu à delegacia de polícia para ser ouvido(a) sobre os fatos, conforme tabela 5.

Tabela 5: Profissionais de saúde ouvidas(os) em sede policial no âmbito dos autos de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Profissional de saúde foi ouvida(o) sobre os fatos no âmbito do inquérito policial	*Casos*	*Proporção*
Sim	16	37,2%
Não	27	62,8%
Total	*43*	*100%*

Fonte: As autoras (2021).

Nos 12 casos em que a mulher foi denunciada pelo Ministério Público à Justiça[12], em sete (58,3%) deles, ao menos um(a) profissional de saúde estava arrolado(a) como testemunha da acusação, e em 10 (83,33%) o conteúdo da denúncia trazia informações constantes do prontuário entregue pelo estabelecimento de saúde, conforme tabelas 6 e 7.

Tabela 6: Profissionais de saúde arroladas(os) como testemunhas de acusação no âmbito dos autos de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Profissional de saúde foi arrolada(o) como testemunha de acusação pelo MP	*Casos*	*Proporção*
Sim	*7*	58,3%
Não	4	41,7%
Total	*12*	*100%*

Fonte: As autoras (2021).

Tabela 7: Conteúdo da denúncia traz informações constantes do prontuário médico entregue pelo estabelecimento de saúde à Polícia Civil, no âmbito dos autos de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Conteúdo da denúncia traz informações do prontuário médico entregue pelo estabelecimento de saúde à Polícia Civil	*Casos*	*Proporção*
Sim	10	83,33%
Não	2	16,67%
Total	*12*	*100%*

Fonte: As autoras (2021).

12 Do total de 43 casos, o Ministério Público promoveu o arquivamento de 29 inquéritos policiais, denunciou 12 mulheres, e dois casos ainda estavam em fase de investigação à época da conclusão da pesquisa.

Do total de casos denunciados, convém mencionar que a 10 mulheres foi oferecida a suspensão condicional do processo; um caso se encontrava na fase de resposta à acusação e em outro a mulher havia sido pronunciada e aguardava julgamento pelo Tribunal do Júri. Quando se buscou analisar, entre as 12 mulheres denunciadas, quantas foram aquelas que tiveram o sigilo do seu prontuário médico violado por profissionais de saúde – seja por meio de comunicação feita à polícia por profissional de saúde, pelo fato de tal profissional ter sido ouvida(o) durante a fase de investigação ou em juízo, ou pela direção da instituição de saúde ter compartilhado o prontuário médico da mulher sem o seu consentimento –, e qual foi o desfecho processual de seu caso, o cenário é o seguinte, conforme tabela 8.

Tabela 8: Relação entre a violação, ou não, do sigilo do prontuário médico por profissionais de saúde e o desfecho processual, no âmbito dos autos de mulheres criminalizadas por aborto no estado do Paraná (2017 a 2019):

Caso	Houve violação do sigilo do prontuário médico?	Desfecho Processual
1	Sim	SCP
2	Sim	SCP
3	Sim	SCP
4	Sim	SCP
5	Sim	SCP
6	Sim	RA
7	Não	SCP
8	Sim	SCP
9	Sim	RD
10	Sim	I
11	Não	SCP
12	Sim	P

Fonte: As autoras (2021).

Legenda: SCP – suspensão condicional do processo; RA – resposta à acusação (em tramitação à época da conclusão da pesquisa); RD – rejeição da denúncia pelo Juízo; I – impronunciada (neste caso, a mulher foi pronunciada, mas a decisão foi reformada em 2ª instância); P – pronunciada (à época da conclusão da pesquisa, a mulher aguardava julgamento pelo Tribunal do Júri).

Diante desse quadro, observa-se que, no Paraná, a situação não difere da encontrada em outros estados[13]: realizar um procedimento abortivo

13 Sobre pesquisas empíricas que demonstram a contribuição do sistema de saúde para a criminalização de mulheres por aborto, ver Ardaillon, 2000; Gonçalves; Lapa, 2008; Cunha; Noronha; Vestena, 2012; Rio de

inseguro e fora das hipóteses legais, e posteriormente procurar atendimento médico para evitar sequelas físicas e emocionais ou até mesmo a própria morte, expõe as mulheres ao risco da criminalização. Aquelas e aqueles a quem cabe cuidar das mulheres em uma situação de risco não raro as denunciam a policiais militares ou a policiais civis, por meio de ligações para as delegacias, e até mesmo permitem que as(os) policiais levem consigo os prontuários sigilosos e que interroguem informalmente a mulher ainda sob efeito de medicamentos, por vezes sangrando, sobre macas de hospitais.

Para salvar suas vidas ou ao menos evitar agravos à sua saúde[14], as mulheres relatam suas histórias e admitem o uso de medicamentos abortivos, submetem-se a exames de urina, de gravidez e de sangue, a ultrassons de suas cavidades uterinas, mamas e abdômen – informações que passam a fazer parte do conteúdo de seu prontuário médico e que depois são enviadas prontamente a delegadas(os) de polícia, incriminando as pacientes.

Por fim, tais profissionais de saúde são ouvidas(os) pela polícia, por vezes comparecem espontaneamente à delegacia de polícia para relatar os fatos com os prontuários em mãos, são arroladas(os) como testemunhas e acusação na denúncia oferecida à Justiça pelo Ministério Público, e, em juízo, expõem informações que ouviram à beira do leito hospitalar, enquanto exerciam a função de cuidadoras(es).

Em um cenário parecido com o encontrado em outros achados de pesquisa, tais mulheres são, em sua maioria, usuárias do Sistema Único de Saúde (SUS); ou seja, profissionais concursadas(os) ou contratadas(os) pelo Estado para atender a população, em sua maioria pobre, negra e sem acesso à Justiça[15], reforçam a seletividade penal dos processos de criminalização

Janeiro, 2018; Melo, 2018; São Paulo, 2018; Ribeiro, 2020. A respeito da discussão dogmática sobre a (in) efetividade da criminalização do aborto, ver Benincasa, 2019; Melo, 2019; Melo, 2020.

14 Como afirmam Faúndes e Barcelatto (2004, p. 81), "as mortes de mulheres relacionadas ao aborto são apenas a ponta de um grande *iceberg*", pois, de acordo com Torres (2007, p. 2), "centenas de milhares de mulheres, todos os anos, sofrem terríveis consequências físicas e psíquicas em razão do abortamento realizado em condições precárias e inseguras: infecções que se instalam nas paredes do útero ou que migram para as trompas, para os ovários ou para a cavidade abdominal (doença inflamatória pélvica – DIP); lesões traumáticas ou químicas dos genitais e outros órgãos pélvicos; reações tóxicas a produtos ingeridos ou introduzidos nos genitais; hemorragias, que acarretam anemia, choque e morte ou que exigem transfusões sanguíneas de emergência, que as expõem a altos riscos de peritonite e contaminação com HIV e outras infecções; septicemia e choque séptico; retirada das trompas, dos ovários e do útero; obstrução das trompas, que pode conduzi-las à esterilidade ou à gravidez tubária, outra causa dramática de morte materna; dores pélvicas crônicas; limitação da vida diária e das atividades sexuais; e depressão e complicações psicológicas em situações de pressão".

15 De acordo com a Pesquisa Nacional de Saúde (PNS) de 2019, realizada pelo Instituto Brasileiro de Geografia e Estatística, 69,9% dos usuários do SUS são mulheres, 60,9% são pretos ou pardos e 64,7% dos usuários têm rendimento domiciliar *per capita* de até 1 salário-mínimo. Ainda, a pesquisa apontou que os usuários

no Brasil, em que pesem as normativas internacional, interamericana e nacional expressamente vedarem tal prática, como se verá a seguir.

Ainda, observam-se as relações de poder que atravessam os encontros entre tais mulheres e a equipe de saúde, e que devem ser lidas também por um filtro de gênero, já que sua vulnerabilidade decorre tanto de sua condição de pacientes como de sua condição de mulheres[16].

4 O ABORTO FORA DAS HIPÓTESES LEGAIS, A PROVA PENAL E O SIGILO MÉDICO SOB A PERSPECTIVA DE GÊNERO

Como se pode depreender da normativa abordada em tópico anterior, a violação do sigilo do prontuário médico é crime previsto no Código Penal e afronta a Constituição Federal de 1988, a qual, para dar efetividade à garantia fundamental de tutela da intimidade e da vida privada, previu igualmente a proteção ao sigilo da fonte no exercício de algumas profissões, notadamente as profissões de saúde, como deixam claro os seus códigos de ética, que abordam expressamente essa problemática.

O fato de que a instrução probatória não pode se lastrear em informações produzidas mediante depoimento de pessoas que, por dever profissional, não podem compartilhá-las nem mesmo com o Estado nos permite afirmar, igualmente, que, caso tais provas sejam produzidas, devem ser consideradas não somente ilegítimas, mas ilícitas, já que ferem não apenas normas de direito processual, mas de direito material (Grinover, 1976).

Logo, a violação do sigilo do prontuário médico de mulheres que abortam clandestinamente e confessam o ato às equipes de saúde – violação que ocorre seja por meio da comunicação do fato à polícia, seja pela entrega do prontuário médico à(ao) delegada(o) de polícia ou à(ao) Magistrada(o), seja pelo depoimento de tais profissionais em juízo ou ouvidas(os) ainda em sede policial – deve ensejar o arquivamento do inquérito policial ou o trancamento da ação penal, não podendo, igualmente, embasar decisão judicial, sob pena de nulidade absoluta da sentença.

---

com rendimento entre 1 e 3 salários-mínimos são 32,4% do total, enquanto somente 2,9% deles estão acima dessa faixa de renda; por fim, 53,8% não tinham ocupação profissional à época da pesquisa.

16 A respeito das relações de poder que se estabelecem entre as equipes de saúde e a mulher que procura atendimento médico no contexto do abortamento, ver Dios (2016). Sobre a relação entre médico(a) e paciente, perpassada pela produção de documentos que buscam inquirir sobre a verdade a respeito do corpo e do sofrimento, sugerimos ver Fassin e D'Halluin (2005).

Para além da normativa jurídica já citada, tal trabalho buscou identificar e responder à pergunta se haveria, igualmente, uma normativa específica que abordasse a questão da tutela da intimidade e da vida privada e o direito ao sigilo dos dados do prontuário médico nos casos de aborto praticado fora das hipóteses legais.

Identificou-se que o próprio Estado brasileiro tem, ao longo dos últimos 15 a 20 anos, regulamentado a temática por meio de Normas Técnicas elaboradas pelo Ministério da Saúde, o qual, pelo menos até recentemente, vinha abordando o aborto clandestino e inseguro como um problema de saúde pública[17], considerado como uma das principais causas das altas taxas de mortalidade materna no país. Determina a última edição da Norma Técnica "Atenção Humanizada ao Abortamento":

> Ética Profissional
>
> Do sigilo profissional
>
> Diante de abortamento espontâneo ou provocado, o(a) médico(a) ou qualquer profissional de saúde não pode comunicar o fato à autoridade policial, judicial, nem ao Ministério Público, pois o sigilo na prática profissional da assistência à saúde é um dever legal e ético, salvo para proteção da usuária e com o seu consentimento. O não cumprimento da norma legal pode ensejar procedimento criminal, civil e ético-profissional contra quem revelou a informação, respondendo por todos os danos causados à mulher. (Brasil, 2014, p. 19)

Outros documentos produzidos pelo Estado brasileiro que demonstram a importância de se respeitar o sigilo nesse contexto dizem respeito aos sete relatórios enviados pelo Brasil ao Comitê que fiscaliza o cumprimento da Convenção para a Eliminação de Todas as Formas de Discriminação Contra a Mulher (CEDAW)[18], o chamado Comitê CEDAW. O campo da saúde sexual e reprodutiva tem sido considerado o que impõe os maiores desafios ao Brasil para a efetivação dos direitos humanos das mulheres, de acordo com o Comitê, baseado em informações que lhes foram enviadas

17 A obra *Aborto e saúde pública no Brasil – 20 anos*, publicada em 2009 pelo Ministério da Saúde, é um exemplo de como o País enfrentava a temática, tratando-o como um problema de saúde pública e inclusive custeando pesquisas de grande impacto e extensão sobre o aborto desde um ponto de vista da saúde pública e da saúde reprodutiva das mulheres (Brasil, 2009).

18 A CEDAW foi promulgada pelo Brasil em 13 de setembro de 2002, via Decreto nº 4.377 do Congresso Nacional. Portanto, o Brasil, ao ratificar a CEDAW, adota-a como legislação doméstica e reconhece a jurisdição internacional dos órgãos responsáveis por estabelecer os mecanismos de monitoramento e proteção dos direitos humanos das mulheres.

pelo próprio governo federal entre os anos de 2003 a 2012, referentes a estatísticas que abrangem o período de 1985 a 2011.

Após o envio dos sete relatórios pelo Brasil, em resposta, o Comitê tem apontado a criminalização do aborto como um aspecto importante a contribuir para o alto índice de mortalidade materna no País[19], uma vez que a escolha política pela criminalização da conduta contribui para o alto número de abortos clandestinos e inseguros, e para a falta de acesso das mulheres à assistência médica célere, humanizada e integral. Devido a esse cenário, o Comitê CEDAW recomendou ao Estado brasileiro que garanta o acesso das mulheres a serviços de saúde de qualidade para a gestão de complicações decorrentes de abortos inseguros, o que envolve, entre outras ações, respeitar o sigilo das informações do prontuário médico e não comunicar os casos ao SJC.

Ainda em decorrência de tais práticas, que são um problema não apenas no Brasil, mas em vários países que optaram por criminalizar o aborto, o Comitê expediu a "Recomendação Geral nº 24 – As mulheres e a saúde" aos Estados-partes para que respeitem o direito das pacientes à confidencialidade no atendimento no âmbito da saúde:

> d) Embora a falta de respeito pela confidencialidade dos pacientes afete tanto os homens como as mulheres, poderá dissuadir as mulheres de procurarem aconselhamento e tratamento e, por conseguinte, afetar negativamente a sua saúde e bem-estar. Por esta razão, as mulheres estão menos dispostas a procurarem cuidados médicos para tratamento de doenças do trato genital, para a contracepção ou *para os abortos incompletos* e em casos em que tenham sofrido violência física e sexual. (CEDAW, 1999, p. 3 – grifos nossos)

Não obstante tais documentos determinarem aos Estados-partes a mudança de postura, na prática se observa que as recomendações da CEDAW ao Brasil são pouco cumpridas e possuem pouca eficácia no campo da proteção à saúde da mulher e à maternidade, e da redução de danos no âmbito dos abortos clandestinos e inseguros (Barroso; Pinto; Andrade, 2020). Com efeito, como conclui Sciammarella, a positivação dos direitos humanos e a incorporação dos tratados internacionais no ordenamento jurídico interno não tiveram o poder de "influir na norma penal que criminalizou o aborto

19 Importante mencionar que, nos anos de 2003, 2007 e 2012, anos em que houve manifestação do Comitê CEDAW sobre o cenário brasileiro, observou-se o aumento do número de óbitos maternos no Brasil, ao invés de redução (Barroso; Pinto; Andrade, 2020, p. 12).

ou na interpretação dessas normas pelos tribunais à luz do direito internacional dos direitos das mulheres" (2010, p. 2).

Mesmo diante da normalização social de tal prática entre profissionais de saúde e, por outro lado, da vedação legal, constitucional e convencional que a acompanha, ela é objeto de poucas reflexões jurídico-doutrinárias no Brasil.

Uma pesquisa bibliográfica exploratória por nós realizada na Plataforma da Revista dos Tribunais *Online*[20] em março de 2021, a fim de identificar a produção doutrinária a respeito do uso das informações do prontuário médico para a criminalização de mulheres por aborto, trouxe-nos resultados inexpressivos. Os descritores utilizados no campo de busca "doutrina" foram: "aborto e sigilo médico"; "aborto e segredo médico"; "aborto e segredo profissional"; "aborto e sigilo profissional"; "aborto e prova ilícita" e "aborto e prova penal".

Do total de 82 referências encontradas, 64 foram excluídas por não se adequarem ao objetivo proposto e 10 foram excluídas por se tratarem de artigos repetidos (presentes nos resultados da aplicação de mais de um descritor ou por aparecerem duplicados nos resultados da aplicação de um mesmo descritor). No total, 8 (oito) artigos[21], publicados nos anos de 1939, 1960, 1990, 1994, 1998, 2007, 2015 e 2019, tratavam do tema envolvendo o aborto autoprovocado, a questão do sigilo médico e a prova penal. Após análise de conteúdo (Bardin, 2011), observou-se que três eram favoráveis à violação do sigilo do prontuário médico nos casos de aborto induzido pela própria paciente[22], dois se colocaram de forma contrária[23] e, nos demais três artigos analisados, não se identificou posicionamento a respeito[24].

---

20 A base de dados da Revista dos Tribunais *online* foi escolhida por se constituir em uma importante fonte secundária de informações para a pesquisa jurídica, com indexação de 37 revistas (fontes primárias da pesquisa jurídica), tratando-se, portanto, de uma das maiores bases de dados sobre pesquisa jurídica no País.

21 As obras analisadas foram *Segredo profissional* (1939, republicado em 2010), de José Duarte; *A responsabilidade civil do médico* (1990), de Miguel Kfouri Neto; *Aborto: a polêmica interrupção voluntária ou necessária da gravidez – Uma questão criminal ou de saúde pública?* (1994), de Dagma Paulino dos Reis; *Do valor do consentimento no abortamento criminoso* (1998), de Vicente de Paulo Vicente de Azevedo; *Aborto inseguro: é necessário reduzir riscos* (2007), de José Henrique Rodrigues Torres; *O segredo médico e as informações à polícia e à Justiça* (1960; republicado em 2010), de A. Almeida Júnior; *Considerações sobre o aborto: o conflituoso enfoque penal e ético – Questão relevante de saúde pública* (2015), de Elias Farah; e *Sigilo médico em psiquiatria e psiquiatria forense* (2015), de Thiago Fernando da Silva, Elias Abdalla Filho e Gustavo Bonini Castellana.

22 Azevedo (1998); Duarte (2010); Kfouri Neto (1990).

23 Almeida Júnior (2010); Torres (2007).

24 Silva; Abdalla Filho; Castellana (2019); Farah (2015); Reis (1994).

Em relação à pesquisa documental em autos de mulheres criminalizadas por aborto no estado do Paraná e em tramitação entre 2017 e 2019, observou-se que, igualmente, das 12 mulheres denunciadas pelo Ministério Público, em apenas dois casos a defesa abordou a questão da violação do sigilo dos dados do prontuário médico, seja ainda durante a investigação policial, seja já na fase processual.

Em um primeiro caso[25], o advogado constituído pela mulher a instruiu para que enviasse comunicado por escrito ao hospital que a denunciou à polícia por aborto determinando que apenas com sua autorização por escrito o prontuário médico poderia ser enviado à Polícia Civil para fins de investigação do suposto crime. Assim, em um primeiro momento, o hospital, diante de requisição do delegado de polícia para que concedesse acesso ao documento, afirmou que não poderia atender ao pedido pela ausência de autorização por parte da paciente investigada.

Diante da recusa do hospital, o delegado de polícia ajuizou cautelar inominada criminal requisitando ao juiz do Tribunal do Júri da comarca que determinasse a juntada do prontuário aos autos, pedido que então foi atendido pelo hospital, com o envio do documento dias depois. Nesse caso, a defesa da mulher não se insurgiu contra a decisão concedida no âmbito do inquérito policial, tampouco impetrou *habeas corpus* para trancar a ação penal quando da denúncia e posterior oferecimento da suspensão condicional do processo pelo Ministério Público, que sua cliente aceitou e cumpriu em dois anos, com o consequente arquivamento do feito.

No outro caso analisado[26], a advogada dativa nomeada pelo juízo já para a audiência de suspensão condicional do processo informou em sede de resposta à acusação que sua cliente, diante das condições injustas impostas pelo Ministério Público, reservava-se o direito de primeiramente responder à acusação, em busca de 1) uma reavaliação do juízo quanto à sua decisão de receber a denúncia, rejeitando-a por ausência de justa causa, e 2) subsidiariamente, da absolvição sumária de sua cliente; apenas posteriormente, caso nenhum dos pedidos fosse deferido pelo juízo, defendia a tese de que, somente então, caberia ao MP propor a SCP.

25 Por critérios éticos, optou-se por não descrever o número dos autos, a fim de que não fosse identificada a ré, uma vez que os casos relatados ocorreram em cidades pequenas do interior do estado.

26 Ver nota de rodapé acima.

Na resposta à acusação, a advogada argumentou a respeito do uso não consentido das informações da paciente constantes em seu prontuário médico, além da oitiva do médico "delator" em sede policial, condutas que embasaram o indiciamento e a posterior denúncia da sua cliente à Justiça, como critério para a inadmissibilidade da prova (por se tratar de prova ilícita). À época da conclusão deste trabalho, os autos estavam conclusos para decisão do juízo, não sendo possível vislumbrar se o argumento foi ou não acolhido pela Magistrada para absolver sumariamente a acusada ou decidir pelo não recebimento da denúncia.

Nos dois casos, analisados na pesquisa documental, em que a defesa, em algum momento da investigação policial ou do processo judicial, abordou tal problemática, em nenhum houve o acionamento de um discurso que permitisse questionar e problematizar, sob uma perspectiva de gênero, de que forma tal prática violava direitos específicos das mulheres.

Percebe-se, portanto, um apagamento/uma neutralização da violência de gênero perpetrada no âmbito dos processos de criminalização do aborto, da experiência das mulheres e do corpo sexuado que é alvo de políticas de criminalização que buscam impor a maternidade de forma compulsória, violando, assim, os direitos sexuais e reprodutivos da população feminina, em geral pobre e negra, que precisa "escolher" entre se autoincriminar perante a equipe de saúde ou morrer/suportar em seu corpo sequelas graves de natureza física e emocional.

Não foram mencionados, pela Defesa, os tratados e convenções internacionais de proteção dos direitos humanos das mulheres ou normativas do Ministério da Saúde que visam a humanizar o atendimento em situações de abortamento clandestino e inseguro, tampouco a necessidade de se descriminalizar o aborto no Brasil e de se adotar uma perspectiva de gênero na investigação, processamento e julgamento de mulheres por aborto autoprovocado.

Em relação aos artigos pesquisados, como já mencionado anteriormente, de um total de oito, apenas dois se posicionaram contrários à violação do sigilo do prontuário com fins de criminalização de mulheres por aborto induzido; no entanto, apenas um expressamente abordou a temática desde uma perspectiva de gênero.

Percebe-se que essa grave violação de direitos é pouco discutida pela doutrina e alvo de poucas produções intelectuais no campo das Ciências Criminais, ao mesmo tempo em que aquelas e aqueles que operam o SJC

não trazem para os autos discussões que problematizem tais práticas e que permitam que a discussão avance para os tribunais superiores. A Defesa raramente aborda o uso dos prontuários médicos de suas clientes para a criminalização delas por aborto como uma violação de direitos fundamentais – e, quando o faz, não traz para a arena jurídica as particularidades de gênero envolvidas, o que também não desafia promotoras(es) e juízas(es) a enfrentarem a questão.

Conclui-se, portanto, que há pouca doutrina a respeito do tema, o que permitiria aos operadores do SJC construir e utilizar argumentos que levem em conta a experiência das mulheres que abortam e são posteriormente criminalizadas, e demonstrar como certas práticas judiciárias operam de forma discriminatória e violam os direitos humanos da população feminina brasileira.

No que diz respeito, por fim, às implicações da violação do sigilo do prontuário médico no pós-abortamento clandestino e inseguro para os direitos humanos das mulheres, há vasta literatura e também documentos oficiais que demonstram seus efeitos deletérios, os quais serão explorados adiante.

5 AS IMPLICAÇÕES DA VIOLAÇÃO DO SIGILO DO PRONTUÁRIO MÉDICO NO PÓS-ABORTAMENTO CLANDESTINO E INSEGURO PARA OS DIREITOS HUMANOS DAS MULHERES

A criminalização do aborto no Brasil, em que pese nem todas as mulheres serem responsabilizadas penalmente pelo ato, enseja riscos à saúde e à vida das mulheres que não podem ser menosprezados. De acordo com estatísticas do Ministério da Saúde, o aborto inseguro foi a quarta causa de mortalidade materna no País de 1996 a 2018, com 1.896 casos notificados (Brasil, 2020).

No entanto, como apontam Faúndes e Barcelatto na nota de rodapé nº 12 deste trabalho, as mortes são a ponta do *iceberg*, uma vez que mesmo as mulheres que sobrevivem ao procedimento enfrentam inúmeros problemas de saúde, de ordem física e psicológica, e milhares de internações são registradas via Sistema Único de Saúde para tratar de complicações decorrentes de um aborto realizado em condições clandestinas e, portanto, inseguras.

Além disso, estudos têm demonstrado que as mulheres, cientes de que a prática da comunicação dos casos ao SJC é comum nos estabeleci-

mentos de saúde, e que, não raro, seus dados médicos serão compartilhados com a polícia, optam por não procurar ajuda, ou a procuram tardiamente (Carvalho; Paes, 2014; Brasil, 2018), e tendem a omitir às equipes de saúde a informação de que induziram um aborto. Tal omissão pode prejudicar o diagnóstico célere e preciso e, portanto, o sucesso do tratamento das complicações no pós-abortamento clandestino e inseguro, assim como também pode dificultar a produção e interpretação de estatísticas sobre a amplitude do fenômeno do aborto clandestino e inseguro no País.

Igualmente, estudos como o de Parpinelli (2000), Valongueiro (2000) e Carvalho *et al*. (2008) têm apontado o impacto de tais decisões – tanto da mulher por não relatar o aborto à(ao) profissional, quanto desta(e) por não registrar o aborto em prontuário justamente por temer que os dados venham a ser usados como meio de prova contra sua paciente – para a subnotificação dos casos de mortalidade materna cuja causa é o aborto clandestino e inseguro. Tal realidade é corroborada por Nota Técnica produzida por profissionais do próprio Ministério da Saúde e apresentada na audiência pública realizada em 2018 pelo Supremo Tribunal Federal (STF) referente à Arguição de Descumprimento de Preceito Fundamental (ADPF) nº 442, que discute a descriminalização do aborto no Brasil:

> Vale destacar o grande desafio que é reduzir a mortalidade materna por abortamento em países onde o aborto se realiza na clandestinidade e ilegalidade. A ilegalidade aumenta a chance de complicação, pois leva as mulheres a não declararem ter interrompido a gestação quando são atendidas na emergência dos hospitais, dificultando o diagnóstico e intervenção médica oportuna, agravando o risco de morte. (Brasil, 2018, p. 7)

Assim, é possível afirmar que o ato de reportar a paciente à polícia, compartilhar dados do seu prontuário médico com o SJC e depor contra essa paciente em juízo, quando passa a ser de conhecimento público entre as mulheres de uma determinada localidade, enseja comportamentos como os acima descritos, isto é, as mulheres optam por não procurar ajuda médica, ou retardam a busca por um pronto-atendimento, colocando sua saúde e a própria vida em risco.

Por outro lado, quando procuram atendimento, muitas optam por omitir o fato à equipe de saúde – ou as(os) próprias(os) profissionais de saúde, em atitude de respeito aos preceitos éticos da profissão, receosas(os) de que outra(o) profissional venha a violar o sigilo de tais dados, optam por informar no prontuário que o aborto foi espontâneo ou não especificado,

comprometendo as estatísticas sobre a mortalidade materna e sobre a própria dimensão da prática do aborto clandestino e inseguro no Brasil.

Neste sentido, além do direito fundamental à intimidade e à vida privada – mencionado desde logo devido ao íntimo diálogo entre o direito constitucional à intimidade e à vida privada e a garantia do sigilo profissional –, é possível concluir que o direito fundamental à saúde de tais mulheres também resta comprometido ou é totalmente inviabilizado por tais práticas, na maioria das vezes sob omissão do próprio Estado, já que a maior parte de tais violações ocorre no âmbito do sistema público de saúde, afetando mulheres vulnerabilizadas em razão do gênero, da classe e da raça/etnia. E mais: fere-se o princípio constitucional da não discriminação entre homens e mulheres, uma vez que às mulheres são impostas condicionantes e restrições que não são impostas aos homens para o pleno exercício de seu direito à saúde[27].

Igualmente, o princípio da não autoincriminação é ferido nesse contexto (Andrade, 1992; Choukr, 1995; Dias Neto, 1997; Queijo, 2003), já que os dados constantes no prontuário médico são de natureza privada, e, quando são compartilhados com profissionais de saúde, o são em decorrência de uma necessidade – no contexto aqui analisado, para salvar a própria vida ou evitar sequelas graves.

No momento em que, para protegerem sua vida ou integridade corporal, as próprias pacientes revelam o crime, não se trata de confissão voluntária que esteja amparada pelos ditames legais das declarações de pessoa imputada. Mesmo se optam por nada dizer à equipe de saúde, permanecendo caladas, quando aceitam que em seu próprio corpo sejam realizados exames que mais tarde possam vir a demonstrar a prática de um crime, as pacientes tomam tais decisões com fins específicos, que não podem ser desvirtuados a fim de que, posteriormente, tais informações sejam utilizadas com fins de persecução penal contra si mesmas.

Embora o compartilhamento de tais dados seja fruto de uma necessidade (praticamente um caso de *compartilhar ou morrer*), há de se garantir o direito à autodeterminação (Westin, 1970) mesmo nesse contexto difícil em

27 Recentemente, têm sido noticiados casos de estabelecimentos e profissionais de saúde condenados a indenizar pacientes denunciadas à polícia em tais circunstâncias, como um caso ocorrido em 2017, quando uma médica de uma Santa Casa em Araçatuba delatou sua paciente à polícia por suspeita de autoaborto. O hospital foi condenado, em 2021, a indenizar a mulher em R$ 10 mil (Bergamo, 2021).

que o espaço para o exercício da autonomia se encontra mitigado, já que se trata de dados pertencentes à esfera privada de tais mulheres e, por pertencerem a essa esfera, cabe somente às suas titulares a decisão de com quem, quando, para quais fins e em qual medida serão compartilhados.

Conforme relatório divulgado a respeito da violação do sigilo dos dados médicos no contexto do pós-abortamento clandestino e inseguro na América Latina, tal prática tem se disseminado, e vários países têm aprovado leis que chancelam tal violação[28], inclusive com a condicionante de que a mulher só possa receber atendimento após confessar o ato, independentemente de o aborto ter sido espontâneo ou provocado (IPAS, 2016).

> *Cuando las mujeres y niñas son forzadas a confesar haber tenido un aborto ilegal, esto constituye una violación de su derecho a guardar silencio y a estar libre de autoincriminación. Las confesiones obtenidas durante los cuidados de emergencia no deben ser consideradas como evidencia admisible válida para el procesamiento, ya que fueron obtenidas en violación del derecho al debido proceso, y es irrespetuoso del secreto médico y del derecho de la paciente a la privacidad.* (IPAS, 2016, p. 8)

Percebe-se, portanto, que vários direitos fundamentais e, portanto, indisponíveis estão sob ataque diante da prática reiterada da violação, por parte das(os) próprias(os) profissionais de saúde, dos dados do prontuário médico da mulher que induz um aborto, para fins de criminalização. Ao mesmo tempo, observa-se que o tema ainda não é abordado com frequência nem recebe a atenção que lhe é devida por parte da doutrina processual penal, sendo que nem mesmo a defesa das mulheres criminalizadas, ao menos no contexto da pesquisa documental realizada, constrói e aciona argumentos que busquem ver declarada a nulidade das provas baseadas em tais condutas.

Essas atitudes das(os) profissionais de saúde são reconhecidas pelo próprio Estado brasileiro, seja por meio de Normas e Notas Técnicas pro-

28 Apesar da aprovação de leis cada vez mais severas para mulheres incriminadas por aborto, inclusive impondo a profissionais de saúde o ônus de denunciar tais mulheres, convém mencionar rapidamente que a Corte Interamericana de Direitos Humanos (CADH) decidiu que o(a) profissional de saúde tem o dever de respeitar o sigilo mesmo quando a(o) paciente é suspeita(o) ou comprovadamente praticou atos criminosos, o que lhe garante igualmente o direito de não denunciar a(o) paciente e não ser punido por tal ato. Trata-se do caso *De La Cruz Flores vs. Peru* (2004), no qual a CIDH entendeu que o Peru violou o art. 9º da CADH (princípio da legalidade) por "[...] penalizar o ato médico, que não é somente um ato essencialmente lícito, mas também um dever do médico prestá-lo; e por impor aos médicos a obrigação de denunciar possíveis condutas delitivas de seus pacientes com base na informação que obtenham no exercício de sua profissão" (CADH, 2004).

duzidas pelo Ministério da Saúde, seja por meio de relatórios enviados ao Comitê CEDAW, como obstáculos à erradicação dos altos índices de mortalidade materna no País, ao acesso pleno à saúde e, por consequência, à efetivação dos direitos humanos das mulheres. Esse cenário, no entanto, não tem sido objeto de ponderações pela doutrina jurídica e também não é problematizado no âmbito da defesa processual de mulheres criminalizadas por aborto.

CONSIDERAÇÕES FINAIS

O aborto é uma prática que faz parte da vida reprodutiva das mulheres brasileiras, uma vez que uma em cada cinco já realizou tal procedimento até os seus 40 anos de idade. Todos os anos, milhares de mulheres procuram o sistema público de saúde para tratar dos agravos à saúde resultantes da prática clandestina e insegura, e, durante ou logo após esse atendimento, parte delas é delatada pelas(os) próprias(os) profissionais de saúde às autoridades policiais, o que, de acordo com a normativa jurídica nacional e internacional, é prática que viola o sigilo dos dados do prontuário médico da paciente e afronta os seus direitos fundamentais e humanos.

Pesquisas realizadas em várias unidades da federação demonstram que essa é a principal forma de captura da mulher que aborta de forma clandestina pelas teias do SJC, e, no âmbito deste trabalho, o cenário não é distinto, uma vez que 44,18% das mulheres criminalizadas por aborto no Estado do Paraná, cujos autos estavam tramitando entre 2017 e 2019, foram denunciadas à polícia por médicas(os), assistentes sociais e enfermeiras(os); ainda, em 65,11% dos casos, o prontuário médico da mulher foi compartilhado com a Polícia Civil sem a sua autorização; em 37,2% dos autos, ao menos um(a) profissional de saúde foi ouvido(a) na delegacia de polícia sobre os fatos e, em 58,3% dos casos, tais profissionais foram arroladas(os) como testemunha de acusação pelo Ministério Público.

Os dados também nos permitem afirmar que a maioria dessas violações ocorre durante ou após atendimento prestado no sistema público de saúde, sendo que foram identificadas situações em que a mulher é interrogada informalmente por policiais à beira do leito, logo após passarem por curetagem uterina, ou mesmo enquanto ainda aguardam o primeiro atendimento, sob dores intensas e hemorragias.

Dessa análise, também observamos que a violação dos dados constantes do prontuário médico não é questionada pela defesa das mulheres

criminalizadas – à exceção de dois casos –, e que em nenhum caso, mesmo naqueles que de alguma forma mencionaram a problemática, tal conduta foi classificada como uma violação dos direitos fundamentais de suas clientes à intimidade, à vida privada e à saúde, e também aos princípios da não discriminação entre homens e mulheres e da não autoincriminação.

Embora haja normativas jurídicas nacionais e internacionais a respeito da importância de se respeitar o sigilo das informações médicas – e embora tal prática constitua crime –, quando o sigilo é violado no âmbito da criminalização de mulheres por aborto, a temática ainda padece de invisibilidade também junto à produção doutrinária nacional no campo das Ciências Criminais.

Tal cenário demonstra, como já abordado pela Criminologia Feminista, que os processos de criminalização de mulheres sofrem de invisibilidade ou generalização junto à produção do campo – mesmo que levem à morte, majoritariamente, de mulheres jovens, pobres, negras e de baixa escolaridade, e que tal criminalização seja reforçada justamente por aquelas(es) que deveriam lhes prestar cuidados.

REFERÊNCIAS

ALMEIDA JÚNIOR, Antônio A. O segredo médico e as informações à polícia e à justiça. *Revista dos Tribunais*, São Paulo, v. 301, p. 41-49.

ANDRADE, Manoel da Costa. *Sobre a proibição de prova em processo penal*. Coimbra: Coimbra Editora, 1992.

ANDRADE, Mariana Dionísio de. Eficácia das Recomendações da CEDAW e as políticas públicas de proteção à maternidade e saúde da mulher no Brasil. *Revista Eletrônica do Curso de Direito da UFSM*, v. 15, n. 1, p. 1-34, 2020. Disponível em: https://bit.ly/3sDDOOY. Acesso em: 21 ago. 2021.

ARDAILLON, Danielle. Para uma cidadania de corpo inteiro: a insustentável ilicitude do aborto. *Anais do XII Encontro Nacional de Estudos Populacionais*. Associação Brasileira de Estudos Populacionais, p. 1-29, 2000. Disponível em: https://bit.ly/3CQUqqv. Acesso em: 18 ago. 2021.

AZEVEDO, Vicente de Paulo Vicente. Do valor do consentimento no abortamento criminoso. *Revista dos Tribunais*, v. 750, p. 761-777, abr. 1998.

BARCELATTO, José; FAÚNDES, Aníbal. *O drama do aborto*: em busca de um consenso. Campinas: Komedi, 2004.

BARDIN, Laurence. *Análise de conteúdo*. Trad. Luís Antero Reto e Augusto Pinheiro. São Paulo: Edições 70, 2011.

BARROSO, Ana Beatriz de Mendonça; PINTO, Eduardo Régis Girão de Castro; Eficácia das recomendações da CEDAW e as políticas públicas de proteção à maternidade e saúde da mulher no Brasil. *Revista Eletrônica do Curso de Direito da UFSM*, v. 15, n.1, 2020. Disponível em: https://bit.ly/3GuXwlQ Acesso em 21 ago. 2021.

BENINCASA, Camila Danielle de Jesus. *A descriminalização do aborto*: uma análise da partir da criminologia feminista. 2019. 132f. Dissertação (Mestrado em Humanidades, Direitos e Outras Legitimidades) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2019.

BERGAMO, Mônica. Hospital quebra sigilo médico e é condenado a pagar R$ 10 mil a paciente que denunciou por autoaborto. *Jornal Folha de S. Paulo*, São Paulo, 31 out. 2021. Disponível em: https://bit.ly/3DdmYv6. Acesso em: 10 nov. 2021.

BRASIL. *Aborto e saúde pública no Brasil* – 20 anos. Brasília: Ministério da Saúde, 2009. Disponível em: https://bit.ly/3CekJWT. Acesso em: 10 nov. 2021.

_____. *Atenção humanizada ao abortamento* (Norma Técnica). 2. ed. Brasília: Ministério da Saúde, 2011. Disponível em: https://bit.ly/31TjmNw. Acesso em: 21 ago. 2021.

_____. *Constituição da República Federativa Brasileira de 1988*. São Paulo: Saraiva, 2012.

_____. *Decreto-Lei nº 2.848, de 7 de dezembro de 1940* (Código Penal Brasileiro). Rio de Janeiro: Diário Oficial da União, 1940.

_____. *Decreto-Lei nº 3.688, de 3 de outubro de 1941* – Lei das Contravenções Penais. Rio de Janeiro: Diário Oficial da União, 1941.

_____. *Decreto-Lei nº 3.689, de 3 de outubro de 1941* – Código de Processo Penal Brasileiro. Rio de Janeiro: Diário Oficial da União, 1941.

_____. *Interrupção voluntária da gravidez e impacto na saúde da mulher*. Brasília: Ministério da Saúde, 2018. Disponível em: https://bit.ly/3ipjJob. Acesso em: 20 ago. 2021.

_____. *Lei nº 8.069, de 13 de julho de 1990* – Estatuto da Criança e do Adolescente. Brasília: Diário Oficial da União, 16 jul. 1990, p. 13563.

_____. *Lei nº 10.741, de 1º de outubro de 2003* – Dispõe sobre o Estatuto do Idoso e dá outras providências. Brasília: Diário Oficial da União, 3 out. 2003, p. 1.

_____. *Lei nº 13.105, de 16 de março de 2015* (Código de Processo Civil). Brasília: Diário Oficial da União, 17 mar. 2015, p. 1.

_____. *Lei nº 13.146, de 6 de julho de 2015* – Institui a Lei Brasileira da Pessoa com Deficiência (Estatuto da Pessoa com Deficiência). Brasília: Diário Oficial da União, 7 jul. 2015, p. 2.

_____. *Lei nº 13.709, de 14 de agosto de 2018* – Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília: Diário Oficial da União, 15 ago. 2018, p. 59.

_____. *Resolução CFM nº 1.638, de 10 de julho de 2002* – Define prontuário médico e torna obrigatória a criação da Comissão de Revisão de Prontuários nas instituições de saúde. Diário Oficial da União, Brasília, Edição nº 153, Seção 1, p. 184-5, 9 ago. 2002.

_____. *Resolução CFM nº 2.217, de 27 de setembro de 2018* – Aprova o Código de Ética Médica. Diário Oficial da União, Brasília, Edição nº 211, Seção 1, p. 179, 1º nov. 2018.

CAMPOS, Carmen Hein de. *Criminologia feminista*: teoria feminista e crítica às criminologias. Rio de Janeiro: Lumen Juris, 2017.

CARVALHO, Marta Lúcia de Oliveira *et al*. Os registros em prontuário de mulheres atendidas por aborto em um Hospital Universitário de cidade da região Sul do Brasil de 2001 a 2005. *Fazendo Gênero 8* – Corpo, Violência e Poder, Florianópolis, 25 a 28 de agosto de 2008.

CARVALHO, Simone Mendes; PAES, Graciele Oroski. As experiências de mulheres jovens no processo do aborto clandestino – Uma abordagem sociológica. *Revista de Saúde e Sociologia*, São Paulo, v. 23, n. 2, p.548-557, 2014. Disponível em: https://bit.ly/2VXS5uk. Acesso em: 18 ago. 2021.

CHOUKR, Fauzi Hassan. *Garantias constitucionais na investigação criminal*. São Paulo: RT, 1995.

CORTE INTERAMERICANA DE DIREITOS HUMANOS. Caso De La Cruz Flores vs. Peru. Washington: CIDH, 2005. Disponível em: https://bit.ly/3jwm7xS. Acesso em: 30 ago. 2021.

CUNHA, José Ricardo; NORONHA, Rodolfo. VESTENA, Carolina Alves. Mulheres incriminadas por aborto no Tribunal de Justiça do Rio de Janeiro: personagens, discursos e argumentos. In: CARVALHO, Paulo de Barros Ribas (Org.). *Desafios da Constituição*: democracia e Estado no século XXI. Rio de Janeiro: Universidade Federal do Rio de Janeiro, p. 209-224, 2012.

DIAS NETO, Theodomiro. O direito ao silêncio: tratamento nos direitos alemão e norte-americano. *Revista Brasileira de Ciências Criminais*, v. 19, p. 179-204, jul./set. 1997.

DINIZ, Débora; MEDEIROS, Marcelo; MADEIRO, Alberto. Pesquisa Nacional de Aborto 2016. *Revista Pesquisa & Saúde Coletiva* [*online*], Rio de Janeiro, v. 22, n. 2, p. 65-660, 2017. Disponível em: https://bit.ly/39RvIu2. Acesso em: 7 ago. 2021.

DINIZ, Débora; MEDEIROS, Marcelo. Itinerários e métodos do aborto ilegal em cinco capitais brasileiras. *Revista Ciência & Saúde Coletiva*, v. 17, n. 7, p. 1671-1681, 2012. Disponível em: https://bit.ly/3k6mPBi. Acesso em: 18 ago. 2021.

DIOS, Vanessa Canabarro. *A palavra da mulher*: práticas de produção de verdade nos serviços de aborto legal no Brasil. Tese de Doutorado. Universidade de Brasília, 2016. Disponível em: https://bit.ly/35FO5Rx. Acesso em: 7 nov. 2021.

DUARTE, José. Sigilo profissional. *Revista dos Tribunais*, n. 120, jul. 1939.

FARAH, Elias. Considerações sobre o aborto: o conflituoso enfoque penal e ético – Questão relevante de saúde pública. *Revista de Direito de Família e das Sucessões*, v. 4, p. 53-85, abr./jun. 2015.

FASSIN, Didier; D'HALLUIN, Estelle. The truth from the body: medical certificates as ultimate evidence for asylum seekers. *Am Anthropol*, v. 107, n. 4, p. 597-608, 2005.

FUNDAÇÃO PERSEU ABRAMO. *Mulheres brasileiras e gênero nos espaços público e privado*. São Paulo: Fundação Perseu Abramo; Serviço Social do Comércio, agosto de 2020. Disponível em: https://bit.ly/3z4Q2mw. Acesso em: 18 ago. 2021.

GRINOVER, Ada Pelegrini. *Liberdades públicas e processo penal* – As interceptações telefônicas. São Paulo: Saraiva, 1976.

GONÇALVES, Tamara Amoroso; LAPA, Thaís de Souza (Coord.). *Aborto e religião nos tribunais brasileiros*. São Paulo: Instituto para a Promoção da Equidade, 2008. Disponível em: https://bit.ly/3mb23Dd. Acesso em: 18 ago. 2021.

HASSEMER, Winfried. *Introdução aos fundamentos do direito penal*. Trad. Pablo Rodrigo Alflen da Silva. Porto Alegre: Sergio Antonio Fabris, 2005.

IPAS. *Delatando a las mujeres*: el deber de cada prestador/a de servicios de denunciar implicaciones jurídicas y de derechos humanos para los servicios de salud reproductiva en Latinoamérica. Chapel Hill: IPAS, 2016. Disponível em: https://bit.ly/3yumaii. Acesso em: 30 ago. 2021.

KFOURI NETO, Miguel. A responsabilidade civil do médico. *Revista dos Tribunais*, v. 654, p. 57-76, 1990.

MELO, Cínthia Carvalho de. *A (in)eficácia positivo-normativa do crime de aborto provocado pela gestante*: um estudo a partir das decisões dos tribunais superiores e de Minas Gerais. 2020. 151f. Dissertação (Mestrado em Direito) – Programa de Pós-Graduação em Direito da Faculdade de Direito do Sul de Minas (FDSM), Pouso Alegre, 2020.

MELO, Mônica de. *Direito fundamental à vida e ao aborto a partir de uma perspectiva constitucional, de gênero e da criminologia*. 2018. 188f. Tese (Doutorado em Direito Constitucional) – Programa de Pós Graduação em Direito da Pontifícia Universidade Católica de São Paulo (PUC-SP), São Paulo, 2018.

MENDES, Soraia da Rosa. *Criminologia feminista*: novos paradigmas. 2. ed. São Paulo: Saraiva, 2017.

_____. *Processo penal feminista*. São Paulo: Atlas, 2020.

NAFFINE, Ngaire. *Feminism and criminology*. Cambridge: Polity Press, 1997.

OBSERVATÓRIO DE GÊNERO DO GOVERNO DO BRASIL. *VI Relatório Nacional Brasileiro* – Convenção para a Eliminação de Todas as Formas de Discriminação

contra as Mulheres (CEDAW) das Organização das Nações Unidas. Brasília: Secretaria Especial de Políticas para as Mulheres, 2008. 98 p. Disponível em: https://bit.ly/3iGN4xv. Acesso em: 24 ago. 2021.

OLIVEIRA, Jorge Alcibíades Perrone de. Sigilo ou segredo médico – A Ética e o Direito. *Revista de Bioética*, n. 2, v. 9, p. 141-148, 2001. Disponível em: https://bit.ly/3uxaklW. Acesso em: 7 ago. 2021.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS. *Recomendação Geral nº 24* – As mulheres e a saúde. Nova York: Comitê para a sobre a Eliminação de Todas as Formas de Discriminação contra as Mulheres, 1999. Disponível em: https://bit.ly/30i3QKs. Acesso em: 24 ago. 2021.

PARPINELLI, Mary Angela *et al*. Subnotificação da mortalidade materna em Campinas: 1992 a 1994. *Revista Brasileira de Ginecologia e Obstetrícia*, v. 22, n. 1, p. 27-32, jan./fev. 2000. Disponível em: https://bit.ly/2W1s1xS. Acesso em: 24 ago. 2021.

QUEIJO, Maria Elizabeth. *O direito de não produzir prova contra si mesmo*: o princípio "*nemo tenetur se detegere*" e suas decorrências no processo penal. São Paulo: Saraiva, 2003.

REGINATO, Andrea Depieri. Uma introdução à pesquisa documental. In: MACHADO, Maíra Rocha (Org.). *Pesquisar empiricamente o Direito*. São Paulo: Rede de Estudos Empíricos em Direito, p. 189-224, 2017.

REIS, Dagma Paulino. Aborto: a polêmica interrupção voluntária ou necessária da gravidez – Uma questão criminal ou de saúde pública? *Revista dos Tribunais*, v. 709, p. 277-284, nov. 1994.

RIBEIRO, Isabela Lopes Leite. *Mulheres acusadas do crime de aborto*: um estudo dos processos judiciais de 2017 e 2018 no Distrito Federal. 2019. 87f. Dissertação (Mestrado em Direito) – Programa de Pós Graduação em Direito da Universidade de Brasília, Brasília, 2019. Disponível em: https://bit.ly/3aBFgJc. Acesso em: 18 ago. 2021.

RIO DE JANEIRO (Estado). *Entre a morte e a prisão*: quem são as mulheres criminalizadas pela prática de aborto no Rio de Janeiro. Rio de Janeiro: Defensoria Pública Geral do Estado do Rio de Janeiro, 2018, 224 p. Disponível em: https://bit.ly/36Ge8Yg. Acesso em: 18 ago. 2021.

SÃO PAULO (Estado). *30 habeas corpus*: a vida e o processo de mulheres acusadas da prática de aborto em São Paulo. São Paulo: Defensoria Pública do Estado de São Paulo, 2018. 20 p. Disponível em: https://bit.ly/3tiA6Ko. Acesso em: 18 ago. 2021.

SCIAMMARELLA, Ana Paula de Oliveira. Criminosas ou vítimas? Documentação das violações de Direitos Humanos das mulheres criminalizadas. *Fazendo Gênero 9* – Diásporas, Diversidades, Deslocamentos, Florianópolis, 23 a 26 de agosto de 2010.

SILVA, Paulo Eduardo Alves da. Pesquisa em processos judiciais. In: MACHADO, Maíra Rocha (Org.). *Pesquisar empiricamente o Direito*. São Paulo: Rede de Estudos Empíricos em Direito, p. 275-320, 2017.

SILVA, Thiago Fernando da; ABDALLA FILHO, Elias; CASTELLANA, Gustavo Bonini. Sigilo médico em psiquiatria e psiquiatria forense. *Revista de Direito e Medicina*, v. 3, jul./set. 2019.

SMART, Carol. *Women, crime and criminology*: a feminist critique. London: Routledge & Kegan Paul Books, 1976.

TORRES, José Henrique Rodrigues. Aborto inseguro: é necessário reduzir riscos. *Revista Brasileira de Ciências Criminais*, v. 68, p. 27-68, set./out. 2007.

VALONGUEIRO, Sandra. *Mortalidade materna por aborto*: fontes, métodos e instrumentos de estimação. Anais do XII Encontro Nacional de Estudos Populacionais. Disponível em: https://bit.ly/3metb4k. Acesso em: 18 ago. 2021.

WESTIN, Alan. *Privacy and freedom*. New York: Atheneum, 1970.

YIN, Robert K. *Pesquisa qualitativa do início ao fim*. Trad. Daniel Bueno; revisão técnica: Dirceu da Silva. Porto Alegre: Penso, 2016.

Sobre as autoras:

Katie Silene Cáceres Arguello | *E-mail:* arguello.katie@gmail.com

Doutora em Direito e Sociologia pelo Departamento de Anthropologie et Sociologie du Politique – Université Paris 8 – Vincennes-Saint-Denis (2000). Mestre em Direito pela Universidade Federal de Santa Catarina (1994). Professora da Graduação e da Pós-Graduação em Direito da UFPR. Integrante do Instituto de Criminologia e Política Criminal (ICPC/PR). Associada do Instituto Carioca de Criminologia (ICC/RJ). Coordenadora do Núcleo de Criminologia e Política Criminal (PPGD/UFPR) e do Grupo de Estudos em Criminologia Crítica (CNPq).

Vanessa Fogaça Prateano | *E-mail:* vanessa.prateano@gmail.com

Mestranda em Direito do Estado – Área de Concentração Estado, Poder e Controle pelo Programa de Pós-Graduação em Direito (PPGD) da Universidade Federal do Paraná (UFPR). Especialista em Direito Penal e Direito Processual Penal pela Academia Brasileira de Direito Constitucional. Bacharela em Direito – Habilitação em Teoria do Direito e Direitos Humanos pela UFPR (2019). Bacharela em Comunicação Social – Habilitação em Jornalismo pela UFPR (2010). Pesquisadora associada do Núcleo de Criminologia e Política Criminal do PPGD-UFPR e da Rede de Estudos Empíricos em Direito (REED). Assessora Jurídica do Núcleo de Promoção e Defesa dos Direitos da Mulher (NUDEM) da Defensoria Pública do Estado do Paraná.

Data de submissão: 28 de setembro de 2021.
Data de aceite: 2 de dezembro de 2021.

Protodefesa à Brasileira: Contraditório e Ampla Defesa em Investigações Sigilosas

The Brazilian Protodefense: Adversarialism and the Right to a Fair Trial in Secret Investigations

NATHALIE FRAGOSO[1]
Universidade de São Paulo (USP).

GABRIEL BREZINSKI RODRIGUES[2]
Universidade Estadual do Rio de Janeiro (UERJ).

RESUMO: O emprego de métodos ocultos de investigação sem qualquer contrapeso defensivo permitiu um avanço desmedido dos órgãos de persecução penal sobre direitos fundamentais. Uma concepção democrática de processo penal, no entanto, não é harmonizável com intervenções em direitos protegidos, sem que um contraditório mínimo seja assegurado. Neste sentido, a criação da figura do "Protodefensor" para defender os interesses do acusado, mantendo sigilo em relação a este, parece-nos a melhor forma de restabelecer o equilíbrio processual e evitar afetações desproporcionais de direitos fundamentais. A Defensoria Pública, argumentamos, seria a instituição brasileira mais adequada para exercer este papel. Logo, ainda que uma lei regulamentadora deste novo papel processual seja necessária, parece-nos, desde logo, compatível com a Constituição a "Protodefensoria Pública".

PALAVRAS-CHAVE: Métodos ocultos; processo penal; protodefesa; prova.

ABSTRACT: The use of covert investigative methods in the absence of any defensive checks and balances has allowed an unrestrained advance of the organs of criminal prosecution over fundamental rights. A democratic criminal procedure, however, cannot be harmonized with interventions on protected rights, without ensuring an adversarial procedure. In this sense, the creation of the "Protodefender", who maintains confidentiality in relation to the accused, but safeguards his interests,

1 Orcid: https://orcid.org/0000-0001-6514-007X.
2 Orcid: http://orcid.org/0000-0001-5494-5569.

seems the best way to reestablish the procedural balance and avoid disproportionate interference with fundamental rights. The Public Defender's Office, we argue, would be the most appropriate Brazilian institution to perform this role. Therefore, even though a law regulating this new procedural role is necessary, the "Public Protodefender" is fully compatible with the Constitution.

KEYWORDS: Hidden methods; criminal procedure; protodefense; evidence.

INTRODUÇÃO

Já há uma década que Manoel da Costa Andrade expressa inquietação com a expansão das medidas investigativas sigilosas, especialmente a colheita de elementos informacionais que pressupõem o desconhecimento da medida pelo afetado. Dentre tais métodos, encontra-se o que chamamos no Brasil de quebras de sigilos de dados e interceptações telefônicas e telemáticas.

São dois os eixos de preocupações Andrade (2011, p. 536). O primeiro deles é a institucionalização dos métodos ocultos, já que, nos últimos anos, diversos ordenamentos jurídicos legitimaram material e procedimentalmente as então obscuras técnicas de espionagem. O segundo é a massificação dos expedientes invasivos, que, além de terem potencialidade para atingir os direitos fundamentais de múltiplos indivíduos, sejam estes investigados ou não[3], tendem a expandir no mesmo ritmo das inovações tecnológicas, sobretudo as do campo informático (Gomes, 2019, p. 36).

As preocupações aventadas por Andrade são pertinentes. Já é possível perceber que os métodos ocultos de investigação provocaram fissuras na estrutura acusatória do processo penal, sobretudo no esgarçamento das garantias individuais, na erosão de direitos fundamentais clássicos, como as inviolabilidades do sigilo de dados e comunicações, e no total abandono da concepção do acusado enquanto sujeito processual titular de direitos.

3 É digno de nota que o Anteprojeto e Lei de Proteção de Dados para segurança pública e investigação criminal limite o acesso a dados pessoais sigilosos de pessoas investigadas, quando controlados por pessoas jurídicas de direito privado. Nesse caso, o acesso dependerá de ordem judicial baseada em indícios de envolvimento dos titulares de dados afetados em infração penal e na demonstração de necessidade dos dados à investigação.

Lendo Andrade, Malan (2016, p. 220) também estabelece correlação entre a popularização dos métodos ocultos com a policialização da investigação, com conseguinte hipertrofia dos poderes da polícia judiciária e privatização de certos aspectos da investigação preliminar, já que, por vezes, o particular acaba por atuar como *longa manus* dos órgãos de repressão[4]. Para mais, o autor aponta que a naturalização de tais processos frequentemente deturpa a natureza retrospectiva do processo penal, que passa a desempenhar uma função de prevenção à futura prática de infrações penais, ao contrário do tradicional papel de reconstrução factual (2021, *livro digital*).

No âmbito do processo penal brasileiro, o impacto mais contundente da expansão dos métodos ocultos de investigação é o abandono quase absoluto do respeito ao contraditório durante a fase pré-processual. Por aqui, alega-se que a persecução só pode ser devidamente construída quando o indivíduo é espionado sem ter conhecimento da vigilância. Aliado ao argumento utilitarista, sustenta-se, no âmbito jurídico-processual, que os elementos colhidos durante a investigação preliminar são apenas informes para estruturação do conjunto probatório que, conforme juízo de conveniência do Ministério Público, serão posteriormente levados à instrução. Consequentemente, durante a fase pré-processual, não se possibilita qualquer tipo de reação do investigado quanto a afetações de seus direitos fundamentais, seja por conta do desconhecimento da medida, seja pela própria concepção inquisitiva de um contraditório diferido ou inexistente.

Diante do panorama narrado, se, de fato, os métodos ocultos são um caminho sem volta (Malan, 2021, *livro digital*), é necessário pensar em filtros de irracionalidade e estruturas de controles viáveis dentro da estrutura constitucional. A lei expressa parece ser a primeira barreira desejável. Porém, no obscuro submundo das cautelares, é preciso que algum contrapeso defensivo atue ainda no âmbito dos procedimentos judiciais que visam à obtenção oculta de dados, sobretudo pelo contraditório não ser oportunizado ao afetado.

A retórica usual sobre o pleito é de que o próprio Ministério Público é capaz de exercer este controle. Contudo, não é incomum que os métodos ocultos de investigação sejam instados, iniciados, requeridos, modulados e executados pelo próprio *Parquet*. Portanto, parece impossível que um mes-

4 Cite-se, como exemplo, o caso dos provedores de aplicações de internet, que, por força do art. 15 do Decreto nº 8.771, de 11 de maio de 2016, devem sintetizar e estruturar os dados requisitados por ordem judicial de forma a facilitar o posterior tratamento pelos órgãos de repressão do Estado.

mo ente ocupe o papel de acusador e de freio da persecução. Assim, na busca por uma solução que contemple um modelo acusatório de processo penal, o presente trabalho parte da proposição feita por Schünemann (2007) do Protodefensor, figura que teria como uma de suas atribuições restabelecer a paridade de armas em procedimentos sigilosos, para então verificar se a instituição brasileira da Defensoria Pública poderia atuar como contrapeso defensivo nos procedimentos ocultos nos quais o investigado não pode ser pessoalmente representado.

1 MÉTODOS OCULTOS E CONTRADITÓRIO NO CENÁRIO PRÉ-PROCESSUAL BRASILEIRO

Na experiência brasileira, a conivência com a proliferação dos meios ocultos de investigação é facilmente perceptível pela jurisprudência processual-penal dos últimos vinte anos, que habitualmente preza pela primazia dos interesses relacionados à tutela do poder punitivo, ainda que em detrimento de direitos fundamentais individuais. Cite-se, como exemplo, os reiterados precedentes que, pautados em alegações de complexidade da investigação, permitem sucessivas renovações de interceptações telefônicas, muito embora o corolário lógico-interpretativo de uma medida restritiva de direitos seja pela interpretação de maior garantia[5]. Ou mesmo a controversa decisão do Superior Tribunal de Justiça ("STJ"), nos autos do RMS 61302/RJ[6], que considerou proporcional a autorização judicial para identificação de um enorme conjunto indeterminado de usuários próximos ao local de um homicídio, aduzindo que a legislação brasileira não exigiria a demonstração da indispensabilidade da medida ou de qualquer elemento de individualização pessoal. Dita decisão demonstra que, mesmo na ausência de previsão

5 A constitucionalidade das renovações sucessivas e, por conseguinte, a interpretação do termo "renovável por igual tempo" descrito no art. 5º da Lei nº 9.296/96 são temas do pendente Recurso Extraordinário nº 625263/PR, interposto pelo Ministério Público Federal e de relatoria do Ministro Gilmar Mendes. Vale notar que, em 13 de junho de 2016, o Supremo Tribunal Federal, por unanimidade, reconheceu a existência de repercussão geral do tema.

6 Trata-se de recurso interposto pelo Google, LLC contra decisão do Tribunal de Justiça do Rio de Janeiro que determinou o fornecimento de registros de conexão e Device IDs de usuários que pesquisaram o nome da vereadora e as combinações "Vereadora Marielle", "Agenda vereadora Marielle", "Casa das Pretas", "Agenda vereadora Marielle" e "Rua dos Inválidos", entre outros termos, nos dias 7 de 14 de março de 2018. A ordem também determinava o fornecimento dos registros de conexão e cruzamento de dados dos dispositivos que passaram pelo pedágio da Transolímpia, via que liga os bairros Recreio dos Bandeirantes e Deodoro, em 2 de dezembro de 2018, entre 11h05 e 11h20. Segundo o Ministério Público do Rio de Janeiro, autor do pedido judicial que deu origem à quebra de sigilo, a quebra genérica destes dados seria a única maneira de identificar os responsáveis pelo homicídio da vereadora Marielle Franco e do motorista Anderson Gomes. A empresa Google, LLC, por outro lado, defende que a medida é desproporcional e ilegal, já que não há norma autorizativa para a quebra de sigilo de dados de uma gama de pessoas não identificadas e sequer individualizadas. Ver: BRASIL. STJ, ROMS 61302/RJ, 2019/0199132-0, 3ª S., Rel. Min. Rogerio Schietti Cruz, DJ 26.08.2020, DJe 04.09.2020.

legal expressa para a prática, tribunais brasileiros recorrem à ponderação de matriz pós-positivista para restringir direitos fundamentais e à analogia à lei ordinária para justificar e legitimar a prova obtida.

Cientes da usual chancela exercida pelo Judiciário, autores como Tavares e Casara (p. 96, 2020), além de Prado (2014, p. 64), sustentam que as informações obtidas durante uma apuração oculta devem ser submetidas em sua totalidade ao contraditório, uma vez que esta é a única forma de possibilitar a reação da parte que não participou da produção. Prado (2014, p. 69), inclusive, defende que todo o processo de obtenção deve ser devidamente registrado e documentado, mantendo-se a cadeia de custódia da prova para que o investigado seja posteriormente capaz de verificar que os dados obtidos não sofreram alteração ou contaminação.

Embora fundamental enquanto filtro do poder punitivo, as diretrizes acima defendidas somente operam efeitos práticos quando o dado colhido furtivamente é submetido à instrução processual penal, já que os métodos ocultos de investigação se desenvolvem ainda no âmbito da investigação preliminar, geralmente em medidas cautelares sigilosas, deferidas por um juiz, mas conduzidas pela polícia judiciária ou pelo aparato de apoio do Ministério Público. Tais autos são os reais detentores dos elementos informativos, pois, na prática, o inquérito é instaurado apenas como formalidade para que a autoridade investigativa possa postular por medidas sigilosas (Malan, 2016, p. 219). Por conseguinte, no âmbito do processo penal brasileiro, a coleta de fontes reais de prova não é submetida a qualquer contraditório, já que o lugar-comum teórico é de que os elementos colhidos não servem diretamente à formação do convencimento do juiz, mas apenas como informe para estruturação do conjunto probatório que posteriormente será levado à instrução. Portanto, se o método oculto de investigação for realizado em completo descompasso com a legislação, sua consequência prática é a inadmissibilidade probatória do elemento colhido, e não a nulidade do ato (Malan, 2021, *livro digital*).

O supracitado sustentáculo teórico confere ao cenário brasileiro das investigações preliminares ocultas duas particularidades. A primeira delas diz respeito à possibilidade quase nula do investigado de conhecer a afetação de direitos, enquanto a segunda relaciona-se com o progressivo abandono da necessidade de lei autorizativa para as medidas, uma vez que, dentre outros constructos argumentativos que serão abordados a seguir, compreende-se que eventuais abusos na coleta de fontes reais de prova poderão ser verificados durante a instrução.

1.1 Opacidade das cautelares: o afetado deve ter ciência da quebra de sigilo de dados?

A estrutura processual das medidas cautelares, em que apenas Polícia, Ministério Público e Juiz participam, tornam o procedimento bastante opaco[7]. É verdade que, após ser instado por diversas vezes, o Supremo Tribunal Federal emitiu a Súmula Vinculante nº 14, garantindo ao defensor o acesso aos elementos de prova. Contudo, dito acesso restringe-se aos elementos de prova já documentados, pois argumenta-se que o conhecimento completo do feito pelo defensor prejudicaria as diligências ocultas em andamento. Em outras palavras, o defensor somente poderá acessar a cautelar quando o órgão que conduz o procedimento finalizar os trabalhos, pois alega-se que a persecução só pode ser devidamente construída quando o indivíduo é espionado sem ter conhecimento da vigilância[8].

Ocorre que, diferente das execuções de medidas cuja manifestação corporal elimina o sigilo (ao exemplo da prisão e a busca e apreensão), a concretização da colheita oculta não pressupõe que o investigado tome ciência do ocorrido. Além disso, a legislação não determina que um investigado seja notificado da existência de um inquérito ou da finalização deste, o que permite que um indivíduo tenha seu sigilo completamente devassado sem jamais tomar consciência do ocorrido, haja vista que não é necessário que a investigação gere uma denúncia ou mesmo, na visão do citado precedente do STJ, que os indivíduos cujo sigilo tenha sido escancarado sejam individualizados como suspeitos. Em fato, ainda que alguém saiba que é formalmente investigado, a nebulosidade burocrática que gravita sobre os

7 Interessante notar que os problemas da parca visibilidade das medidas de interceptação telefônica foram investigados em um passado recente, durante a Comissão Parlamentar de Inquérito sobre escutas telefônicas clandestinas, conhecida como "CPI dos Grampos". O resultado indireto dos trabalhos foi a elaboração, pelo Conselho Nacional de Justiça, da Resolução nº 59, de 09.09.2008, que disciplina e uniformiza as rotinas de interceptações, além de criar o Sistema Nacional de Controle de Interceptações Telefônicas. Embora pouco tenham impactado sobre a curva ascendente das interceptações, ditas estruturas de controle ao menos trouxeram visibilidade quantitativa do fenômeno. Sucede que, até a publicação deste artigo, ainda não há qualquer controle ou gestão, por parte do Judiciário, sobre ordens judiciais que requisitam dados aos provedores de aplicação. Não obstante, sobre o impacto da CPI e da Resolução do CNJ sobre a série histórica das interceptações, vale conferir o trabalho de: SANTORO, Antonio Eduardo Ramires; TAVARES, Natália Lucero Frias; OLIVEIRA, Anderson Affonso de. A interceptação telefônica no contexto dos maxiprocessos no Brasil: uma análise quantitativa e qualitativa dos dados entre 2007 e 2017. *Revista Brasileira de Ciências Criminais*, São Paulo, v. 26, n. 143, p. 89-116, maio 2018.

8 Alguns delegados de polícia entendem, ainda, que seria lícito optar por sequer responder um requerimento defensivo de listagem de procedimentos sigilosos instaurados contra uma pessoa, já que o simples fato de saber que há um procedimento sigiloso instaurado pode mudar o comportamento do investigado. Ver: BARBOSA, Ruchester Marreiros; COSTA, Adriano Sousa. Não responder a *phishing* ou a *firewall* defensivos não é abuso de autoridade. *Conjur – Consultor Jurídico*, 23 de março de 2021. Disponível em: https://www.conjur.com.br/2021-mar-23/academia-policia-nao-responder-phishing-ou-firewall-defensivos-nao-abuso-autoridade. Acesso em: 26 ago. 2021.

procedimentos é suficiente para que, mesmo após a conclusão da medida, o indivíduo sequer tome ciência de que ela ocorreu, especialmente quando a quebra de sigilo não reúne elementos úteis para respaldar a imputação. Por isso, há quem defenda que o Estado deve notificar o afetado tão logo seja encerrada a necessidade do sigilo[9] (Gleizer, 2021, *livro digital*).

No caminho contrário, encontram-se aqueles que pretendem validar a hipertrofia do estado de polícia, argumentando que, quando os dados obtidos são incapazes de formar uma acusação criminal, inexiste razão para que o afetado tenha ciência da quebra, já que o material obtido continuará restrito ao sistema de justiça criminal. Dito coro parece deliberadamente olvidar que tanto o Superior Tribunal de Justiça quando o Supremo Tribunal Federal[10] já permitiram o compartilhamento, para fins de processo administrativo disciplinar, das informações colhidas por interceptação telefônica, desde que o Juízo Criminal tenha expressamente autorizado (Amorim, 2010, p. 13). Portanto, não parece que a potencialidade deletéria dos dados obtidos com a quebra de sigilo encerre-se quando o Ministério Público decide não oferecer denúncia.

A nosso ver, a questão deveria ser analisada sobre a perspectiva do direito à autodeterminação informacional. Tal como reconhecido pelo Tribunal Constitucional Federal Alemão no BVerfGE 65, 1[11], e pelo Supremo Tribunal Federal Brasileiro na ADIn 6387/DF[12], trata-se de um direito fundamental autônomo, com base na proteção constitucional dos dados pessoais, que permite o controle ideal do indivíduo sobre as informações que lhe digam respeito, a fim de concretizar o livre desenvolvimento da persona-

9 É digno de nota que o Anteprojeto de Lei de Proteção de Dados para Segurança Pública e Persecução Penal, produzido pela Comissão de juristas liderada pelo Ministro Néfi Cordeiro, do Superior Tribunal de Justiça, e entregue em 5 de novembro de 2020 ao então Presidente da Câmara dos Deputados Rodrigo Maia (DEM-RJ), veda a proibição genérica de notificação dos titulares de dados quando do fornecimento em resposta a requisição administrativa ou judicial sigilosa. A notificação de afetados está em linha com o contraditório previsto no CPP, por exemplo, para imposição de medida cautelar (art. 282, § 3º). Na ausência de notificação, a defesa de direitos do titular de dados depende da disposição do controlador de dados. Texto integral do Anteprojeto em: https://www2.camara.leg.br/atividade-legislativa/comissoes/grupos-de-trabalho/56a-legislatura/comissao-de-juristas-dados-pessoais-seguranca-publica/documentos/outros-documentos/DADOSAnteprojetocomissaoprotecaodadossegurancapersecucaoFINAL.pdf. Acesso em: 20 nov. 2021.

10 Respectivamente: ROMS 16429, 200300870460, 6ª T., DJe 23.06.2008 e Inq-QO 2.42404/RJ, J. 24.04.2007.

11 No famoso julgamento da Lei do Censo de 1983, o Tribunal Constitucional Federal Alemão reconheceu o direito fundamental à autodeterminação informativa, declarando que "não existem mais dados insignificantes no contexto do processamento eletrônico de dados". Ver: ALEMANHA. Tribunal Constitucional Federal Alemão, BVerfGE – Repertório oficial de jurisprudência, 65, 1, "Recenseamento" (Volkszählung), Reclamação Constitucional contra Ato Normativo, 15.12.1983.

12 BRASIL. STF, ADIn 6387/DF, 0090566-08.2020.1.00.0000, Tribunal Pleno, Relª Rosa Weber, DJ 07.05.2020, publ. 12.11.2020.

lidade. Por meio dele, o cidadão deve poder determinar quem sabe o que sobre ele, como, quando e em que circunstância.

Ainda que, em regra, a autodeterminação informativa esteja inserida no paradigma de abstenção que orienta a proteção de direitos fundamentais, no âmbito do interesse do indivíduo em conhecer as medidas informacionais do Estado que afetem seus direitos, dito direito ultrapassa sua função de defesa e incorpora um aspecto positivo (Gleizer *et al.*, p. 39). Nesse sentido, se o problema da notificação do investigado após a conclusão da colheita oculta for enfrentado a partir desta chave, parece inevitável que, se tais dados seguem armazenados nos bancos de dados da justiça e dos órgãos de persecução, é dever do estado cientificar o titular para que este tenha ciência do tratamento[13], pois o direito do afetado de tomar conhecimento da intervenção informacional tem por fundamento mitigar os prejuízos causados pelo levantamento e processamento de seus dados pessoais ocorridos sem sua ciência e sem o seu consentimento, permitindo, por conseguinte, que o titular possa opor-se a qualquer ilegitimidade (Gleizer *et al.*, p. 148).

1.2 O esgarçamento da legalidade estrita em medidas investigativas ocultas

Voltando à rotina procedimental dos meios de obtenção ocultos, constata-se que, adaptados à estrutura processual das medidas cautelares e cientes de que seu pleito chegará de forma unilateral ao Magistrado, algumas autoridades policiais e representantes do Ministério Público têm avançado interpretações expansivas em seus requerimentos de quebra, especialmente no campo das comunicações armazenadas e dos metadados (Antonialli; Abreu, 2019, p. 63), já que a ausência de leis específicas autorizativas e o ar de inovação no "combate à criminalidade" por meio da informática dão margem para diversos constructos argumentativos. Ditos silogismos, como narram Antonialli e Abreu (2019, p. 63), são frequentemente aceitos pelo Judiciário, seja pela ausência de vozes dissidentes no processo sigiloso, seja por adesão subjetiva do julgador à narrativa. Não é incomum, por exem-

13 Novamente, é digna de nota a solução aventada pelo Anteprojeto de Lei de Proteção de Dados para Segurança Pública e Persecução Penal, produzido pela Comissão de juristas liderada pelo Ministro Néfi Cordeiro, do Superior Tribunal de Justiça, e entreguem, em 5 de novembro de 2020, ao então Presidente da Câmara dos Deputados Rodrigo Maia (DEM-RJ). Consoante o art. 19, I e II, da proposta, o titular terá o direito a obter do controlador, mediante requisição, a confirmação da existência de tratamento e o acesso aos dados. Contudo, o art. 20 permite o adiamento da prestação de informações para evitar prejuízo para a persecução penal, segurança do Estado e para proteger direitos e garantias de terceiros. Não obstante, por força dos parágrafos seguintes, os motivos da recusa ou da limitação do acesso devem restar claros, podendo o titular levar o caso ao Conselho Nacional de Justiça ou iniciar ação judicial.

plo, que diversos anos de conversas armazenadas sejam compreendidos como dados com grau de proteção inferior aos quinze dias prospectivos de uma interceptação telefônica, já que a ausência de norma fundamentadora da intervenção jurisdicional sobre o conteúdo das comunicações privadas, descrita no art. 10, § 2º, do Marco Civil da Internet[14], permitiria, na visão dos órgãos de persecução, que conversas fossem violadas após a ponderação típica das buscas e apreensões (proporcionalidade e razoabilidade), ao revés dos critérios mais rígidos do art. 2º da Lei de Interceptações Telefônicas[15].

Diferentemente do que vem sendo aceito pelos tribunais, as ações estatais de intervenção sobre direitos fundamentais carecem de fundamentação de ordem formal e material (Gleizer, 2021, *livro digital*). A verificação destes fundamentos faz parte do que Orlandino Gleizer denomina de âmbito da justificação, que, no plano dos direitos fundamentais sem reserva de lei, como é o caso do sigilo de dados, deve ser interpretado considerando o direito ao sigilo sob o prisma de garantia de proteção elevada, com maior peso na relação de proporcionalidade (Gleizer *et al.*, 2021, p. 35). Consequentemente, dito sigilo só pode ser restringido para tutelar valores de mesma hierarquia constitucional, ao passo que a intervenção só se torna justificável quando legalmente fundada. Portanto, a autorização legal clara, determinada e específica é pressuposto necessário para execução de qualquer ação interventiva estatal (Gleizer, 2021, *livro digital*), uma vez que o

14 Importante notar que, no âmbito da Constituição Federal, o sigilo das comunicações (art. 5º, XII) é tratado com preponderância, admitindo-se sua violação apenas para fins de investigação criminal ou instrução processual penal. Por conta disso, a possibilidade de violação do conteúdo das comunicações privadas, descrito no art. 10, § 2º, do Marco Civil da Internet, vem sendo compreendida como medida restrita ao processo penal: "Entendo, nessa linha de raciocínio, que a adequada exegese dos arts. 7º, II e III, e 10, § 2º, do Marco Civil da Internet, à luz do art. 5º, XII, da Constituição da República, conduz à conclusão inequívoca de que, à maneira das comunicações telefônicas, a inviolabilidade do sigilo das comunicações realizadas pela internet somente pode ser excepcionada, por ordem judicial, no âmbito da persecução penal. Na expressa dicção da Constituição, 'para fins de investigação criminal ou instrução processual penal'". Trecho do voto da Ministra Rosa Weber, do Supremo Tribunal Federal, prolatado nos autos da ADIn 5.527/DF, 4000753-38.2016.1.00.0000, Relª Min. Rosa Weber.

15 Não obstante a pressão dos órgãos de acusação, alguns tribunais vêm timidamente construindo o entendimento de que, na ausência de norma definidora do art. 10, § 2º, do Marco Civil da Internet, para que uma quebra de sigilos de conversas armazenadas seja proporcional, é necessário que o pedido contenha, por analogia, os requisitos descritos no art. 2º da Lei de Interceptações (Lei nº 9.296/1996), como (i) indícios razoáveis de autoria e participação em infrações penais, (ii) imprescindibilidade da prova (iii) e que o fato investigado constitua infração punida com pena de reclusão. Além de coerente com a dinâmica constitucional descrita no art. 5º, XII, dito entendimento tem o condão de filtrar quebras de sigilo desarrazoadas, como a obtenção de todas as conversas de um indivíduo para apurar o suposto cometimento de uma injúria simples. Ver: HC 315.220/RS, julgado pela Sexta Turma do STJ; RHC 67.379/RN, julgado pela Quinta Turma do STJ; e MS 0101473-97.2020.8.26.9000, MS 5028556-66.2021.8.09.0000 e MS 4020797-90.2017.8.24.0000, julgados, respectivamente, pelos Tribunais de Justiça de São Paulo, Goiás e Santa Catarina.

limite para a persecução penal democraticamente estabelecido por meio da reserva de lei e parlamentar[16] é uma garantia do próprio indivíduo.

Também desmorona frente à dogmática constitucional da proteção de dados o argumento, usualmente empregada para justificar medidas atípicas no processo penal, de que a autorização para a restrição de direitos fundamentais não carece de norma expressa, já que a competência constitucionalmente estabelecida aos órgãos de persecução penal (art. 144, § 4º, da CF) conferiria liberdade suficiente para que o investigador inovasse procedimentalmente, desde que autorizada por ordem judicial. Calha, como pontua Gleizer *et al.* (2021, p. 42-43), que as normas que distribuem e organizam a estrutura do Estado não se confundem com normas autorizativas, pois o fato de o constituinte atribuir competências para a realização de uma tarefa (como apurar infrações) não implica, automaticamente, que tudo mais esteja autorizado para o exercício desta função. Portanto, são incorretas as decisões que pretendem superar a ausência de lei fundamentadora por meio da ponderação sobre a necessidade de uma intervenção no caso concreto, haja vista que ditos pronunciamentos violam os limites impostos aos poderes Executivo e Judiciário, na medida em que não há prévio consentimento democrático sobre o limite e a forma de execução da medida (Gleizer *et al.*, 2021, p. 45).

Entrando no âmbito das intervenções legitimadas por lei, mesmo nos casos em que há norma autorizativa, como no caso das interceptações telefônicas, são frequentes as decisões sustentando que a disposição que fundamenta a medida é uma lei regulamentadora parcialmente limitante, que tão somente estabelece requisitos gerais e um roteiro mínimo não restritivo para a execução da quebra, ao contrário de autorizar e limitar a intervenção penal sob o direito fundamental (Wolter, 2018, p. 40). A erosão dos limites para a atividade persecutória que esta linha de entendimento proporciona já é sentida pelos Tribunais Superiores, que, nos últimos cinco anos, depararam-se com casos-escândalo, demonstrando que a inversão do paradigma "lei limitadora" para "norma reguladora" conferiu liberdade para que os órgãos de persecução inovassem artificiosamente nos meios de obtenção de prova. Nos autos do REsp 1806792/SP[17], p. ex., foi preciso que o STJ

16 A concepção de "reserva parlamentar" parte da ideia de que "as decisões essenciais sobre os pressupostos, as circunstâncias e as consequências das intervenções devem ser tomadas pelo legislador e não podem ser delegadas à administração ou ao tribunal" (Gleizer *et al.*, 2021, p. 43).

17 BRASIL. STJ, REsp 1806792/SP, 2019/0103023-2, 6ª T., Relª Min. Laurita Vaz, DJ 11.05.2021, DJe 25.05.2021.

declarasse a ilegalidade de uma determinação judicial para que uma segunda linha do terminal visado fosse entregue, pela operadora de telefonia, aos promotores do caso, permitindo que estes se substituíssem ao investigado e exercessem controle total do terminal, com conseguinte ingresso no aplicativo de mensageria do investigado. Em outro famoso episódio, foi preciso que o Supremo Tribunal Federal suspendesse a decisão ilegalmente aplicada pelo juízo da 2ª Vara Criminal de Duque de Caxias, que, fundado abstratamente no art. 5º, XI, da Constituição Federal, determinou – ao arrepio da lei – que um provedor de aplicação alterasse seus serviços, na forma sugerida pelo Ministério Público Estadual, removendo a criptografia de ponta-a-ponta para viabilizar a interceptação telefônica, sob pena de suspensão dos serviços (Colli; Lopes Jr., 2016).

Além da ilegalidade declarada pelos Tribunais Superiores, os supracitados casos guardam outra familiaridade. Ambos tratam de medidas requeridas ativamente pelo Ministério Público, no âmbito de uma medida cautelar sigilosa em que a participação defensiva é nula. Logo, o Magistrado que deferiu a medida teve acesso apenas aos argumentos elaborados pela acusação. Outro ponto importante é que como tais procedimentos correm em sigilo: se, porventura, as empresas comandadas não tivessem judicializado a ordem recebida, ninguém, além do *Parquet* e do Magistrado, teria ciência da colheita oculta ilícita.

As experiências acima narradas levantam dúvidas sobre a retórica usual, de que o Ministério Público é capaz de atuar como defensor da correta aplicação da lei nos procedimentos sigilosos de obtenção de prova. Portanto, como muitos dos métodos ocultos de investigação são instados, iniciados, requeridos, modulados e executados pelo próprio *Parquet*, é imperioso questionar se o papel exercido pelo Ministério Público no atual sistema acusatório ainda coaduna com o lugar-comum teórico de que um mesmo ente pode ocupar o papel de acusador e de freio da persecução penal.

2 O MINISTÉRIO PÚBLICO ENQUANTO PARTE NO PROCESSO PENAL

A Constituição Federal de 1988 atribui um papel historicamente inusitado ao Ministério Público. Originalmente, a instituição foi instaurada no Brasil durante a colonização, a partir dos "procuradores do rei" do Direito lusitano (Lima; Busato, 2010, p. 267), com a intenção de cumprir a vocação originária – tradicionalmente remetida à experiência francesa – de braço re-

pressivo do poder estatal[18]. Nos anos que se seguiram, com maior ou menor autonomia, o Ministério Público abraçou sua função primordial de executar a política criminal. Foi somente com o novo pacto constitucional de 1988 que o órgão ganhou nova feição, eis que convertido em instituição igualmente incumbida da defesa da ordem jurídica e dos direitos e interesses sociais e individuais.

No âmbito do sistema de justiça criminal, a Constituição de 1988 atribuiu ao Ministério Público as funções de promover, privativamente, a ação penal pública, exercer o controle externo da atividade policial e requisitar diligências investigatórias e a instauração de inquérito policial, indicados os fundamentos jurídicos de suas manifestações processuais (art. 128, I, VII e VIII). Foram também conferidas ao *Parquet* as garantias típicas da magistratura, como vitaliciedade, inamovibilidade e independência funcional.

Diante das incumbências constitucionais e garantias funcionais, argumenta-se que o Ministério Público exerce, no Estado brasileiro, a dupla função de *custus legis* e de acusador. Curiosamente, embora o pressuposto carnelutiano seja justamente o de que, no processo penal, o Ministério Público é uma parte fabricada para salvaguardar a imparcialidade do Magistrado (Lopes Jr., 2015. p. 151), a corrente doutrinária dominante é de que as duas funções não são mutuamente excludentes durante a atividade no âmbito criminal. Logo, a compreensão dessa linha é a de que o Ministério Público é absolutamente imparcial no exercício da persecução penal, sendo capaz de reunir, de forma isenta e durante toda a atividade persecutória (incluindo a fase pré-processual), os elementos acerca do fato criminoso que se mostrem favoráveis ou prejudiciais ao investigado (Souza, 2017, p. 51).

Interessante notar que algumas das batalhas públicas travadas pela própria instituição parecem destoar da ideia de um órgão que atua como *custus legis* da investigação. No RE 593.727/MG[19], por exemplo, a polícia judiciária advogava por sua exclusividade constitucional na apuração de infrações penais, enquanto o Ministério Público pretendia ter seus poderes de investigação reconhecidos. O resultado, proferido pela Suprema Corte em

18 Na conhecida conferência documentada em "A verdade e as formas jurídicas", Foucault narra a íntima relação entre a formação do Estado moderno e os procuradores dos reis. Representando o Soberano, ditos procuradores eram responsáveis por regular a vida em sociedade, exercer a acusação pública e colocar em prática os processos de acumulação de capital necessários para o fortalecimento do monarca. Ver: FOUCAULT, Michel. *A verdade e as formas jurídicas*. Rio de Janeiro: Nau, 2002. p. 53-78.

19 BRASIL. STF, Recurso Extraordinário Representativo de Controvérsia nº 593727/MG, Tribunal Pleno, Rel. Min. Cezar Peluso, DJ 14.05.2015, publ. 08.09.2015.

2015, foi de que o Ministério Público também dispõe de competência para promover investigações penais, tendo liberdade para instaurar seus procedimentos investigatórios criminais ("PICs"), que hoje são razoavelmente regulados pela Resolução CNMP nº 181/2017[20], conquanto, na prática, carreguem muitas das mazelas inquisitoriais inerentes ao inquérito policial.

Ainda que não contestem a imparcialidade do Ministério Público na apuração de infrações penais, autores preocupados com as garantias individuais, como Moreira (2016), alertam para a possível incompatibilidade entre as funções de investigar e acusar, sobretudo diante da cultura acusatória que permeia a instituição. A crítica é relevante, especialmente se pensada sob o viés da teoria da dissonância cognitiva aplicada ao processo penal, clara em demonstrar que os processos mentais inerentes à atividade de perquirição fazem o indivíduo supervalorizar certas informações em detrimento daquelas que descartam a sua premissa (Ritter, 2016, p. 84-128).

Sem olvidar das contribuições da teoria da dissonância cognitiva sobre a atividade investigatória, a reflexão primordial sobre a atuação do *Parquet* na condução ou requisição judicial dos métodos ocultos de investigação parece residir sobre a concepção autocontraditória de "parte imparcial". Críticas mais brandas ao pressuposto podem ser encontradas naqueles que defendem que a imparcialidade é atributo do Magistrado, cabendo ao *Parquet* somente o dever de conduzir a investigação de forma impessoal (Santin, 2008, p. 231-234). No entanto, embora ditas contribuições não sejam descartáveis, os efeitos práticos deletérios da falaciosa crença na imparcialidade do Ministério Público demandam a desconstrução da ideia de que a instituição atua como *custus legis* no processo penal.

Algo que inicialmente deve restar claro é que as alterações inseridas no Código de Processo Penal pela Lei nº 13.964/2019 deram fim à concepção, bastante pobre em termos argumentativos, de que vigorava uma espécie de sistema "misto" no processo penal brasileiro (Pacelli, 2014, p. 14). Na verdade, como defendia Lopes Jr. (2020, *livro digital*), o processo penal brasileiro sempre foi neoinquisitório, já que a gestão da prova – princípio informador de um sistema processual – estava na mão do juiz durante as fases investigativas e processuais. Ocorre que, frente à expressa definição de uma estrutura acusatória, com conseguinte vedação à iniciativa probatória pelo

20 BRASIL. Conselho Nacional do Ministério Público. Resolução nº 181, de 7 de agosto de 2017 [Dispõe sobre instauração e tramitação do procedimento investigatório criminal a cargo do Ministério Público]. Publicado em Diário Eletrônico do CNMP, Caderno Processual, edição de 08.09.2017.

juiz (art. 3º-A do CPP), concretizou-se, em 2019[21], a intenção constitucional de um modelo acusatório, restando apenas os resquícios da mentalidade inquisitiva, percebidos especialmente na fase pré-processual, na qual o sigilo e a ausência de contraditório ainda permanecem[22]. Não é incomum, entretanto, que estes resquícios transbordem para o processo criminal, já que muitos dos atores do sistema de justiça se limitam a validar o que foi produzido durante o inquérito[23].

Tendo em vista que as alterações legislativas tornaram mais claro o papel do *Parquet* na busca pelo convencimento do juízo e na construção dialética da sentença penal, já não há como negar a vigência de um processo penal de partes, que deve contar com igualdade e equilíbrio no sentido material e simbólico (Casara, 2011). Assim, é preciso desconstruir a concepção de um Ministério Público imparcial, tanto pela semântica contraditória e incompatibilidade ontológica (Casara, 2014, p. 940), já que nenhuma parte com iniciativa probatória deixará de deduzir a hipótese – diametralmente oposta à do adversário – que pretende confirmar pela prova requerida, como pelo resgate da função originária da instituição, que nasce para representar o Estado enquanto adversário do sujeito passivo (Lopes Jr., 2020, *livro digital*).

Mais importante, insistir na construção teórica de imparcialidade, até hoje presente no imaginário de muitos membros da instituição (Casara, 2014, p. 941), é ignorar os efeitos que a brutalidade do sistema penal opera sobre seus atores. Ainda que mascarado pela racionalidade burocrática, o cotidiano do atual sistema penal impele seus atores para a naturalização da barbárie, gerando apatia e morte do juízo de autorreflexão sobre a responsabilidade das suas próprias ações (Boldt; Carvalho, 2017, p. 344). Assim, não é incomum que *Parquets* com longo período de atuação criminal se sintam progressivamente incumbidos da missão de "combate ao crime". Resultados

21 Importante notar que, embora o art. 3º-A esteja vigente no ordenamento jurídico brasileiro, até a submissão deste artigo, a eficácia do dispositivo encontra-se suspensa por força da medida cautelar proferida pelo Ministro do Supremo Tribunal Federal Luiz Fux, na Ação Direta de Inconstitucionalidade nº 6.299, do Distrito Federal.

22 Antes da alteração, Raphael Boldt e Thiago Fabres de Carvalho já alertavam que a pretensão de estruturar um modelo processual acusatório na modernidade esbarra no obstáculo epistemológico da raiz fundadora do processo penal, já que a gênese do instituto é de matriz inquisitorial. Ver: BOLDT, Raphael; CARVALHO, Thiago Fabres de. Para além do processo: epistemologia inquisitória e a ilusão do sistema acusatório na modernidade. *Revista Brasileira de Ciências Criminais*, v. 134, p. 323-349, 2017.

23 Ao exemplo de perguntas comuns no cotidiano forense, geralmente feita ao acusado ou às testemunhas durante a audiência de instrução, que podem ser representadas por: "Os fatos que constam no inquérito policial são verdadeiros?".

práticos destes efeitos já foram publicamente percebidos em meio à chamada "Operação Lava-Jato", em especial no que diz respeito à gestão, pelo próprio Ministério Público, de métodos ocultos de investigação (Marques, 2016, p. 205-227). Também é digno de nota a atuação da instituição no já citado RMS 61302/RJ. Neste caso, ao revés de se portar como fiscal impessoal das garantias constitucionais, *Parquets* do Rio de Janeiro abertamente advogaram pela violação dos direitos de diversos não investigados, em prol da eficiência da persecução penal que por eles mesmo era conduzida.

Por último, persistir na noção do Ministério Público como *custus legis* é ignorar o papel da linguagem como legitimante na cultura processual. Não há como existir um processo penal com igualdade de tratamento e paridade de armas quando o *Parquet* é visto em uma posição de superioridade em relação à defesa (Karam, 2009, p. 403). É impossível manter um verdadeiro contraditório quando uma parte é vista como mais digna de credibilidade e mais comprometida com a verdade. Logo, um processo penal equilibrado não depende só da igualdade probatória, mas também que a imparcialidade do Magistrado não seja contaminada por preconceitos (Karam, 2009, p. 404).

A delimitação do verdadeiro papel do Ministério Público na atual estrutura acusatória do processo penal parece responder à pergunta que deu origem à digressão. Resta claro que um órgão cuja função legal é ser parte deve ser percebido como parte. Por óbvio, como todos os funcionários públicos, o *Parquet* deve atuar sempre dentro da estrita observância da legalidade, jamais requisitando a produção de prova quando não há lei autorizativa e indícios razoáveis da prática de um injusto (comportamento que também se exige dos delegados de polícia). No entanto, como o Ministério Público tem o papel de buscar o convencimento do Magistrado, as medidas requeridas por essa parte não podem ser recebidas pelo juízo enquanto pedidos desinteressados, proporcionais e adequados. Do contrário, continuarão despontando no cotidiano forense métodos flagrantemente infundados de obtenção de prova, como aqueles citados anteriormente que foram postulados pelo próprio Ministério Público, mas anulados pelos Tribunais Superiores. Da mesma forma, frente à dialética probatória atual, não parece exigível atribuir ao *Parquet* o dever de zelar pelo comedimento de seus próprios pedidos, mesmo porque isto seria cognitivamente impossível. Assim, é preciso explorar quem, dentre o desenho democrático constitucional, poderia atuar como contrapeso defensivo nos procedimentos ocultos nos quais o investigado não pode ser pessoalmente representado.

3 PRIMEIRA OU PROTODEFESA DE DIREITOS EM PROCEDIMENTOS SIGILOSOS

A Constituição Federal assegura "aos litigantes, em processo judicial ou administrativo, e aos acusados em geral [...] o contraditório e ampla defesa, com os meios e recursos a ela inerentes" (CF, art. 5º, LV). Entre os "acusados em geral", há hoje pouca dúvida que se encontrem os sujeitos que o Estado investiga e contra quem já pode adotar medidas restritivas, mesmo que não haja denúncia formal.

Embora as disposições constitucionais sejam claras, em se tratando de métodos ocultos de investigação, sustenta-se que o exercício do direito de defesa deva ser relativizado, já que a eficácia da medida está condicionada à ignorância do investigado. Como já esclarecido, isto resulta na condução unilateral da investigação preliminar pela polícia e pelo Ministério Público, sem qualquer participação dos afetados. Quando muito, a dialética entre teses de acusação e defesa ocorre depois que encerrada a colheita de elementos, tendo como objeto discutir a futura admissibilidade dos elementos informativos, num contraditório diferido.

Diante do caráter sigiloso das medidas, é subtraída à defesa a possibilidade de questionar a conformidade do requerimento de quebra com as razões formais e materiais que, em tese, justificariam a intervenção sob direitos fundamentais. Esse desequilíbrio ganha relevância, primeiro porque nem sempre erros e arbitrariedades serão identificados ou poderão ser desfeitos; segundo, porque permite que os órgãos de persecução penal atuem sem um olhar defensivo vigilante, o que, a longo prazo, dissemina e banaliza o uso de métodos de obtenção da prova que exploram novas tecnologias de comunicação e informação[24].

Para prevenir afetações desproporcionais de direitos fundamentais, vêm sendo aventados mecanismos de compensação da natureza oculta das medidas (Gleizer, 2021, *livro digital*) e de restabelecimento da paridade de armas (Schünemann, 2007, p. 232) em procedimentos sigilosos. Referimo-nos ao debate europeu em que foi formulada a proposição de criação do Protodefensor (ou Euro-), instituição que teria como uma de suas atribuições participar do procedimento, quando da impossibilidade de o afetado defender-se pessoalmente, de modo a preservar garantias processuais penais e da defesa de seus interesses aparentes (Schünemann, 2007, p. 248).

24 Segundo dados do CNJ, chegam a 300.000 interceptações telefônicas ao ano (Santoro; Tavares, 2019, p. 116).

Trata-se, no contexto em que foi proposto, de uma instituição nova e adicional, cuja tarefa exclusiva seria a proteção dos interesses da defesa, de modo a constituir uma "protodefesa" ou "defesa preliminar". Financiada pelo Estado, e obrigada a manter sigilo em relação ao acusado em investigações sigilosas, teria como missão salvaguardar os interesses do investigado e, neste sentido, zelar pelo controle das medidas. A proposta aparece em textos que discutem procedimentos de investigação transnacional (Schünemann, 2007), e institutos de justiça negociais (Schünemann, 2013), como mecanismo que visa ao reequilíbrio processual em procedimentos investigativos sigilosos que afetam substancialmente o desfecho do processo.

No âmbito das discussões sobre a reforma do Direito Penal e Processual Penal Europeu, a figura foi batizada de Eurodefensor, instituição incumbida de prestar assistência à defesa em procedimentos transnacionais, e desempenhar a tarefa preventiva de atuar em procedimentos sigilosos, nos quais não pode haver um advogado constituído. Neste último caso, solicitada uma intervenção "transnacional", o Eurodefensor seria notificado para que um "Protodefensor" representasse os direitos aparentes do réu (Schünemann, 2007, p. 249). Como esclarece Schünemann (2007, p. 250), talvez em resposta às reservas de atores como a *European Criminal Bar Association* (ECBA) que criticam a proposta como uma interferência do Estado no livre exercício da profissão[25], a defesa privada não seria substituída por um defensor designado pelo Estado. Ela seria exercida num momento processual de outro modo vedado à representação do investigado, promovendo uma defesa básica dos interesses aparentes do réu, constituindo-se garantia contra a prova ilícita e afetações desproporcionais de direitos fundamentais eventualmente atingidos.

4 A (PROTO-)DEFENSORIA PÚBLICA

Cogitada por Gleizer (2021, *livro digital*) como a instituição brasileira adequada para exercer o papel, a Defensoria Pública tem atribuições constitucionais e legais compatíveis com essa forma de defesa que, referida e limitada a investigações sigilosas, pressupõe a falta de conhecimento do investigado.

25 ECBA. Cornerstones for a draft regulation on the establishment of a European Public Prosecutor's Office ("EPPO") in accordance with art. 86 par. 1-3 TFEU. Disponível em: http://www.ecba.org/extdocserv/201300207_ECBACORNERSTONESONEPPO.pdf.

A Lei Complementar nº 80/1994, no art. 4º, estabelece entre as funções institucionais da Defensoria Pública a defesa dos direitos e interesses individuais, difusos, coletivos e individuais homogêneos; o acompanhamento do inquérito policial, inclusive com a comunicação imediata da prisão em flagrante pela autoridade policial, quando o preso não constituir advogado; a atuação nos estabelecimentos policiais, penitenciários e de internação de adolescentes, visando a assegurar às pessoas, sob quaisquer circunstâncias, o exercício pleno de seus direitos e garantias fundamentais; a preservação e reparação dos direitos de pessoas vítimas de tortura, abusos sexuais, discriminação ou qualquer outra forma de opressão ou violência, propiciando o acompanhamento e o atendimento interdisciplinar das vítimas; o exercício da curadoria especial nos casos previstos em lei (XVI); inclusive contra as pessoas jurídicas de direito público, conforme § 2º do mesmo artigo.

A Constituição, desde a Emenda Constitucional nº 80/2014, atribui-lhe, "como expressão e instrumento do regime democrático", a promoção dos direitos humanos e a defesa, em todos os graus, judicial e extrajudicial, dos direitos individuais e coletivos, de forma integral e gratuita, aos necessitados, na forma do inciso LXXIV do art. 5º. As alterações no texto constitucional e na Lei nº 80/1994 são, aliás, interpretadas de modo a extrapolar as atribuições da instituição do universo de pessoas ou grupos vulneráveis, à defesa de valores constitucionalmente assegurados de maneira geral (Kirchner; Barbosa, 2014).

As atribuições relacionadas ao exercício da defesa criminal efetiva de indivíduos criminalizados abarcam, portanto, o quanto esperado de um "Protodefensor". É ainda possível compreender o art. 72, parágrafo único, do CPC/2015, que prevê a nomeação de curador especial, como lastro para a atuação da Defensoria nos procedimentos em comento.

A curadoria especial é instituto processual de caráter protetivo, destinado a garantir a tutela dos interesses de pessoas cuja situação de vulnerabilidade obsta a ciência do processo e o exercício da defesa (Esteves; Silva, 2015). É nessa qualidade que a Defensoria atua, por exemplo, em substituição ao réu revel preso e réu revel citado por edital ou com hora certa; e atua independente da condição econômica do sujeito.

A vulnerabilidade, nestes casos, não é econômica, mas jurídica. Portanto, no exercício da curatela especial, enquanto defende interesses do réu, a Defensoria Pública exerce a função institucional para proteger valores relevantes do ordenamento (González, 2018, p. 97). Embora o instituto

nos pareça desde logo aplicável por analogia, há especificidades relevantes desta forma de representação, que seriam mais bem endereçadas por lei, ao exemplo do dever de guardar sigilo em relação ao investigado. Ademais, dos termos do CPC e da Lei Complementar nº 80/1994, do art. 4º, XVI, entende-se que a legitimação extraordinária deve ser outorgada expressamente pela legislação processual para viabilizar o contraditório e o devido processo legal.

À afinidade com o instituto da curadoria especial, soma-se às razões para indicação da instituição o trabalho que já vem conduzindo na defesa institucional de garantias. A Defensoria Pública desempenha a defesa de um enorme contingente de pessoas em âmbito penal e, como instituição, reúne condições informacionais[26] e humanas para a incidência estratégica sobre a atuação das agências penais do País, seja no exercício da defesa criminal, seja nessa primeira ou protodefesa. Claro, há desafios inerentes ao encargo operacional que uma proposta dessa magnitude implica, já que a defensoria passaria a atuar em uma fase processual a qual, até então, não lhe era exigível a presença. Ainda assim, com a devida dotação orçamentária, a instituição, cuja missão é garantir a assistência jurídica integral gratuita, judicial e extrajudicial, poderia colocar a proposta em prática.

Por fim, embora não seja propriamente uma controvérsia, cabe indicar que a defesa de direitos de sujeitos, com fundamento outro que a condição econômica, vem sendo admitida no Judiciário. A esse respeito, cabe lembrar o primeiro *habeas corpus*[27] coletivo admitido pelo Supremo Tribunal Federal e as considerações sobre a admissibilidade da ação, que compreenderam a ponderação sobre a legitimidade ativa na modalidade coletiva do *writ*. O relator, Ministro Ricardo Lewandowski, seguido pela maioria da 2ª Turma, considerou necessária a eleição de certos parâmetros de legitimidade, como aliás, comum, no processo em coletivização, e o fez por analogia ao instituto do mandado de injunção coletivo (art. 12, IV, da Lei nº 13.300/2016). A Defensoria Pública teria então legitimidade para a defesa das pacientes, já que se tratava de tutela relevante para a promoção

26 São notáveis, por exemplo, a sistematização de dados sobre audiências de custódia e o estudo sobre sentenças em processos relacionados ao tráfico de drogas no Rio de Janeiro. No primeiro, a DPRJ apresenta uma análise das decisões tomadas em audiência de custódia, como meio de acompanhar o cumprimento do HC 143.641. No último, a Defensoria Pública do Estado do Rio de Janeiro analisou a motivação e critérios de condenação em sentenças judiciais em processos que apuram crimes relacionados ao tráfico de drogas na cidade do Rio de Janeiro e sua Região Metropolitana (Haber, 2018).

27 BRASIL. STF, HC 143641/SP, Rel. Min. Ricardo Lewandowski, DJ 24.10.2018, DJe-228 26.10.2018.

dos direitos humanos e a defesa dos direitos individuais e coletivos dos necessitados, na forma do inciso LXXIV do art. 5º da Constituição Federal. Dentre os legitimados do rol previsto na Lei nº 13.300/2016, foi a Defensoria Pública da União a instituição intimada para assumir o polo ativo da ação, numa ação que apontava a violação de direitos individuais de um contingente indeterminado, porém determinável de pessoas, as gestantes e mães submetidas à prisão cautelar no País.

Desde então, outros *habeas corpus* foram apresentados em favor de adolescentes internados (HC 143988/ES[28]), em favor de pessoas condenadas pelo crime de tráfico privilegiado (HC 596.603/SP[29]), relacionados à pandemia de Covid-19 (HC 575495/MG[30] e HC 570440/DF[31]).

CONSIDERAÇÕES FINAIS

As linhas antecedentes demonstraram que a estrutura processual hoje adotada no Brasil tolera o emprego de métodos ocultos de investigação à revelia de qualquer contrapeso defensivo. A narrativa sobre a experiência nacional comprovou que, no campo da colheita de dados informáticos, os vestígios da mentalidade inquisitiva e a falta de legislação expressa sobre o tema permitiram um avanço desmedido dos órgãos de persecução penal sobre direitos fundamentais.

Se, após as alterações operadas no Código de Processo Penal pela Lei nº 13.964/2019, já não é possível tratar o Ministério Público como um ente que não é parte no processo penal, tem-se claro que o art. 5º, LV, da Constituição Federal só restará verdadeiramente concretizado quando a igualdade de tratamento e a paridade de armas estiverem presentes nos espaços burocráticos em que o Estado exerce o poder punitivo, incluindo o que se conveniou chamar de fase pré-processual. Portanto, não obstante a última década tenha representado um avanço contra a obscuridade do inquérito, uma concepção democrática de processo penal que queira preservar a noção de acusado enquanto sujeito titular de direitos não pode aceitar que alguém sofra diversas intervenções em sua esfera de liberdades sem que um contraditório mínimo lhe seja assegurado.

28 BRASIL. STF, HC 143988/ES, Rel. Min. Edson Fachin, DJ 18.06.2019, DJe-135 21.06.2019.
29 BRASIL. STJ, HC 596603/SP, 2020/0170612-1, Rel. Min. João Otávio de Noronha, DJ 04.08.2020.
30 BRASIL. STJ, HC 575495/MG, 2020/0093487-0, Rel. Min. Sebastião Reis Júnior, DJ 28.04.2020.
31 BRASIL. STJ, HC 570440/DF, 2020/0079174-0, Rel. Min. Antonio Saldanha Palheiro, DJ 06.04.2020.

Neste sentido, a criação da figura do Protodefensor, causídico obrigado a manter sigilo em relação ao acusado, mas cuja missão é salvaguardar os interesses deste e zelar pelo controle das medidas, parece-nos a melhor forma de restabelecer o equilíbrio processual. Além disso, no cenário brasileiro, a figura do Protodefensor poderia evitar afetações desproporcionais de direitos fundamentais, especialmente quando se considera que tribunais têm chancelado interceptações por longos períodos e quebras de sigilo de dados contra um conjunto indeterminado de usuários. Crê-se, por exemplo, que quebras de sigilo de dados desproporcionais ao crime investigado poderiam ser evitadas caso o Protodefensor fosse ouvido antes da análise do pedido pelo magistrado. Além disso, ficaria menos a cargo das empresas comandadas insurgir-se contra situações eminentemente ilegais, como o caso dos promotores que requereram uma linha telefônica para substituir o investigado, já que o Protodefensor poderia opor-se à medida ainda em sede de requerimento. No mais, cumpre notar que, na sistemática processual penal brasileira, não há nenhum empecilho de ordem legal que proíba o juiz incumbido do controle da legalidade da investigação criminal de ouvir um causídico público que esteja obrigado a manter sigilo em relação ao acusado. Logo, eventuais entraves da proposta parecem estar mais ligados à falta de interesse de que um contrapeso defensivo atue na fase pré-processual.

Por fim, a nosso ver, ante o desenho democrático da Carta de 1988, a Defensoria Pública é a instituição brasileira mais adequada para exercer o papel de Protodefensor. Como sustentado, o exercício da protodefesa não apenas coaduna com as atribuições constitucionais da instituição, mas também é coerente com atribuições atípicas, como a curadoria especial. Logo, ainda que uma lei regulamentadora deste novo papel processual seja desejável, parece-nos que já há espaço para a construção de uma "Protodefensoria Pública".

REFERÊNCIAS

AMORIM, Maria Carolina de Melo. Considerações sobre o segredo judicial e as provas colhidas com a quebra de sigilo das comunicações telefônicas. *Boletim IBCCrim*, São Paulo, v. 18, n. 211, p. 13-14, jun. 2010.

ANDRADE, Manuel da Costa. Métodos ocultos de investigação (*Pladoyer* para uma teoria geral). In: *Processo penal* – Constituição e crítica: estudos em homenagem ao Dr. Jacinto Nelson Miranda Coutinho. Rio de Janeiro: Lumen Juris, 2011.

ANTONIALLI, Dennys; ABREU, Jacqueline de Souza. O conto do baú do tesouro: a expansão da vigilância pela evolução e popularização de celulares no Brasil. In: *Direitos fundamentais e processo penal na era digital*: doutrina e prática em debate. São Paulo: InternetLab, v. 2, p. 60-89, 2019. Disponível em: https://congresso.internetlab.org.br/wp-content/uploads/2020/08/Direitos-Fudamentais-e-Processo-Penal-na-era-digital-Volume-2.pdf. Acesso em: 26 ago. 2021.

BARBOSA, Ruchester Marreiros; COSTA, Adriano Sousa. Não responder a *phishing* ou a *firewall* defensivos não é abuso de autoridade. *Conjur*, 23 de março de 2021. Disponível em: https://www.conjur.com.br/2021-mar-23/academia-policia-nao-responder-phishing-ou-firewall-defensivos-nao-abuso-autoridade. Acesso em: 26 ago. 2021.

BOLDT, Raphael; CARVALHO, Thiago Fabres de. Para além do processo: epistemologia inquisitória e a ilusão do sistema acusatório na modernidade. *Revista Brasileira de Ciências Criminais*, v. 134, p. 323-349, 2017.

CASARA, Rubens. Igualdade entre MP e defesa contribui para democracia. *Conjur*, 29 de julho de 2011. Disponível em: https://www.conjur.com.br/2011-jul-29/ministerio-publico-constituicao-busca-espaco-publico-republicano. Acesso em: 27 ago. 2021.

_____. O mito da imparcialidade do Ministério Público no processo penal brasileiro: o desvelamento necessário. In: PEDRINHA, Roberta Duboc; FERNANDES, Márcia Adriana (Org.). *Escritos transdisciplinares de criminologia, direito e processo penal*: homenagem aos mestres Vera Malaguti e Nilo Batista. Rio de Janeiro: Revan, 2014.

COLLI, Maciel; LOPES JUNIOR, Aury. Bloqueio do WhatsApp não resolve nenhum problema da investigação. *Conjur*, 22 de julho de 2016. Disponível em: https://www.conjur.com.br/2016-jul-22/limite-penal-bloqueio-WhatsApp-nao-resolve-nenhum-problema-investigacao. Acesso em: 14 fev. 2018.

ESTEVES, Diogo; SILVA, Franklyn Roger Alves. A curadoria especial no novo Código de Processo Civil. In: SOUSA, José Augusto Garcia. *Repercussões do novo CPC* – Defensoria Pública. Salvador: JusPodivm, 2015.

FOUCAULT, Michel. *A verdade e as formas jurídicas*. Rio de Janeiro: Nau, 2002.

GLEIZER, Orlandino. A dogmática dos métodos ocultos de investigação no processo penal. In*: Direitos fundamentais e processo penal na era digital*: doutrina e prática em debate. São Paulo: InternetLab, v. 4, 2021. Disponível em: https://congresso.internetlab.org.br/wp-content/uploads/2021/08/Direitos-fundamentais-e-processo-penal-na-era-digital-Teoria-e-pra%CC%81tica-em-debate-Vol-4.epub. Acesso em: 26 ago. 2021.

GLEIZER, Orlandino; MONTENEGRO, Lucas; VIANA, Eduardo. *O direito de proteção de dados no processo penal e na segurança pública*. 1. ed. Rio de Janeiro: Marcial Pons, 2021.

GOMES, Felismina Solange. A admissibilidade de métodos ocultos de investigação criminal em processo penal: intervenções nas telecomunicações ou comunicações electrónicas. Contributo para a sua reflexão. 2019. 152 f. Dissertação (Mestrado em Direito e Ciências Jurídicas) – Faculdade de Direito, Universidade de Lisboa, 2019.

GONZÁLEZ, Pedro. Defensoria Pública nos 30 anos de Constituição: uma instituição em transformação. *Revista Publicum*, Edição Comemorativa, Rio de Janeiro, v. 4, p. 85-109, 2018.

KARAM, Maria Lúcia. O direito à defesa e a paridade de armas. In: PRADO, Geraldo; MALAN, Diogo (Coord.). *Processo penal e democracia*. Estudos em homenagem aos 20 anos da Constituição da República de 1988. Rio de Janeiro: Lumen Juris, 2009.

KIRCHNER, Felipe; BARBOSA, Rafael Vinheiro Monteiro. In: ROSENBLATT, Ana et al. *Manual de mediação para a Defensoria Pública*. Brasília: Fundação Universidade de Brasília/FUB, 2014.

LIMA, Ana Maria Bourguignon de; BUSATO, Paulo Cesar. A formação histórica do Ministério Público. Origens do Ministério Público na França, em Portugal e no Brasil. *Revista Justiça e Sistema Criminal: Modernas Tendências do Sistema Criminal*, Curitiba, v. 2, n. 3, p. 245-278, jul./dez. 2010.

LOPES JR., Aury. *Direito processual penal*. 17. ed. [livro digital]. São Paulo: Saraiva Educação, 2020.

LOPES JR., Aury. *Fundamentos do processo penal*: introdução crítica. São Paulo: Saraiva, 2015.

MALAN, Diogo. Notas sobre a investigação e prova da criminalidade econômico-financeira organizada. *Revista Brasileira de Direito Processual Penal*, Porto Alegre, v. 2, n. 1, p. 213-238, 2016.

_____. Métodos ocultos, devido processo e o enfrentamento à criminalidade organizada. In: *Direitos fundamentais e processo penal na era digital*: doutrina e prática em debate. São Paulo: InternetLab, v. 4, 2021. Disponível em: https://congresso.internetlab.org.br/wp-content/uploads/2021/08/Direitos-fundamentais-e-processo-penal-na-era-digital-Teoria-e-pra%CC%81tica-em-debate-Vol-4.epub. Acesso em: 26 ago. 2021.

MARQUES, Leonardo Augusto Marinho. Interceptação telefônica e obscurantismo inquisitório: o que aprende com a Lava Jato? *Revista Brasileira de Ciências Criminais*, São Paulo, v. 24, n. 122, p. 205-227, ago. 2016.

MOREIRA, Rômulo de Andrade. Reforma do Código de Processo Penal: a implementação do sistema acusatório no Brasil – O papel do Ministério Público. Palestra proferida no evento "O papel do Ministério Público no processo penal e o sistema acusatório". Bahia, 29 de agosto de 2016.

PACELLI, Eugênio. *Curso de processo penal*. 18. ed. São Paulo: Atlas S.A., 2014.

PRADO, Geraldo. *Prova penal e controle de sistemas epistêmicos*: a quebra da cadeia de custódia das provas obtidas por métodos ocultos. São Paulo: Marcial Pons, 2014.

RITTER, Ruiz. *Imparcialidade no processo penal*: reflexões a partir da teoria da dissonância cognitiva. 2016. 196 f. Dissertação (Mestrado em Ciências Criminais) – Faculdade de Direito, Pontifícia Universidade Católica do Rio Grande do Sul, 2016.

SANTIN, Valter Foleto. Impessoalidade e imparcialidade do Ministério Público na ação penal. *Justitia*, São Paulo, v. 65, n. 199, p. 231-234, jul./dez. 2008.

SANTORO, A. E. R.; TAVARES, N. L. F. (nov. 2019). Diagnóstico sobre o uso da interceptação telefônica no Brasil: uma análise quantitativa e comparativa dos dados até 2018. *Revista Brasileira de Ciências Criminais*, 161(27), p. 101-130.

SANTORO, Antonio Eduardo Ramires; TAVARES, Natália Lucero Frias; OLIVEIRA, Anderson Affonso de. A interceptação telefônica no contexto dos maxiprocessos no Brasil: uma análise quantitativa e qualitativa dos dados entre 2007 e 2017. *Revista Brasileira de Ciências Criminais*, São Paulo, v. 26, n. 143, p. 89-116, maio 2018.

SCHÜNEMANN, Bernd. Alternative-Project for a European Criminal Law and Procedure. *Crim. Law Forum*, 18, 227-251, 2007. Disponível em: https://doi.org/10.1007/s10609-007-9031-z.

_____. *Estudos de direito penal, direito processual penal e filosofia do Direito*. São Paulo: Marcial Pons, 2013.

_____. Risse im Fundament, Flammen im Gebälk:
Die Strafprozessordnung nach 130 Jahren. *ZIS*, p. 484-493, 2009. Disponível em: http://www.zisonline.com/dat/artikel/2009_10_358.pdf.

SOUZA, Alexander Araujo de. Ainda e sempre a imparcialidade do Ministério Público no processo penal: uma tese decididamente garantista. In: DE BEM, Leonardo Schmitt (Org.). *Estudos de Direito Público 1*: aspectos penais e processuais. Belo Horizonte: D'Plácido, 2018.

TAVARES, Juarez; CASARA, Rubens. *Prova e verdade*. São Paulo: Tirant lo Blanch, 2020.

WOLTERS, Jürgen. *O inviolável e o intocável no direito processual penal*: reflexões sobre dignidade humana, proibições de prova, proteção de dados (e separação informacional de poderes) diante da persecução penal. Trad. e org. Luís Greco. 1. ed. São Paulo: Marcial Pons, 2018.

Sobre a autora e o autor:

Nathalie Fragoso | *E-mail:* nathalie.fragoso@gmail.com

Advogada Criminalista e Pesquisadora. Ex-Coordenadora da área de Privacidade e Vigilância do InternetLab. Doutorado e Graduação em Direito pela Faculdade de Direito da Universidade de São Paulo (USP). Possui o Zertifikat in den Grundzügen des deutschen Rechts e o LLM na Ludwig-Maximilians-Universität München.

Gabriel Brezinski Rodrigues | *E-mail:* gbr@tutamail.com

Mestre (2019) e Doutorando em Direito Penal pela Universidade Estadual do Rio de Janeiro (UERJ). Pós-Graduado em Direito Penal e Processual Penal pela Damásio (2015). Graduado pela Faculdade de Direito de Vitória, FDV (2014). Advogado.

Data de submissão: 30 de setembro de 2021.
Data de aceite: 2 de dezembro de 2021.

O RE 1.055.941: um Pretexto para Explorar Alguns Limites à Transmissão, Distribuição, Comunicação, Transferência e Difusão de Dados Pessoais pelo COAF

HELOISA ESTELITA[1]
Fundação Getúlio Vargas (FGV/SP).

RESUMO: Em dezembro de 2019, o Supremo Tribunal Federal decidiu que o COAF poderia revelar informações sigilosas (financeiras) para as autoridades de persecução penal sem a necessidade de autorização judicial prévia (RE 1.055.941). Essa decisão é utilizada no presente trabalho como pretexto para examinar o tratamento de dados pessoais pelo COAF sob a luz da gramática da proteção de dados pessoais. Analisa-se o tratamento de dados pessoais, sensíveis e sigilosos no âmbito das tarefas desempenhadas pelo COAF, especialmente sob o viés das comunicações que esse órgão faz para autoridades nacionais e internacionais. O exame evidencia que há atividades sendo desempenhadas sem normas autorizativas de intervenção em direitos fundamentais, apontando para a necessidade de reforma legislativa para adequar o tratamento de dados pessoais aos *standards* atuais de proteção desse direito fundamental.

PALAVRAS-CHAVE: Lavagem de dinheiro; unidade de inteligência financeira; proteção de dados; sigilo financeiro.

ABSTRACT: In December 2019, the Brazilian Supreme Court, ruled that the Brazilian Financial Intelligence Unit – FIU (COAF) could share confidential (financial) information to law enforcement authorities without prior judicial authorization (RE 1.055.941). This decision is used in the present paper as the background to review the processing of personal data by COAF under the grammar of personal data protection. The processing of personal, sensitive, and confidential data is analyzed vis-à-vis the tasks performed by COAF, particularly regarding the reports this agency makes to national and international authorities. The examination shows that there are activities being carried out without the authorizing norms for intervention in fundamental rights, which points towards the need for legislative reform to adapt the processing of personal data to the current standards of protection of this fundamental right.

1 Orcid: http://orcid.org/0000-0002-5054-4116.

KEYWORDS: Money laundering; financial intelligence unit; data protection; bank secrecy.

SUMÁRIO: I – Introdução: A matéria discutida no STF no RE 1.055.941; II – As premissas; II.1 Proteção de direitos fundamentais: legalidade e proporcionalidade; II.1.a) Dever de abstenção e exigência de normas autorizativas; II.1.b) Normas autorizativas e normas de competência (atribuição); II.1.c) Proteção de dados pessoais como direito fundamental; II.1.d) Separação informacional; II.1.e) Proteção de dados no Brasil; II.2 Privacidade e proteção de dados no tratamento para as medidas de controle e prevenção da lavagem de capitais; II.2.a) Direito à privacidade, direito à proteção de dados e autodeterminação informacional; II.2.b) Dados pessoais, dados pessoais sensíveis e dados pessoais sigilosos; II.3 O tratamento de dados pessoais pelo COAF; II.3.a) Competências e tarefas atribuídas ao COAF; II.3.b) Dados pessoais; II.3.c) Dados pessoais financeiros sigilosos; III – Consequências; III.1 RIFs de ofício ("disseminação espontânea"); III.2 RIFs a pedido ("disseminação a pedido"); III.3 Intercâmbio internacional entre UIFs; IV – À guisa de conclusão; Referências.

I – INTRODUÇÃO: A MATÉRIA DISCUTIDA NO STF NO RE 1.055.941

Em dezembro de 2019, o Supremo Tribunal Federal (STF), ao decidir, com repercussão geral, as controvérsias a ele submetidas no Recurso Extraordinário nº 1.055.941, estabeleceu (1) ser "constitucional o compartilhamento dos relatórios de inteligência financeira da UIF [...] com os órgãos de persecução penal para fins criminais sem prévia autorização judicial, devendo ser resguardado o sigilo das informações em procedimentos formalmente instaurados e sujeitos a posterior controle jurisdicional"; e (2) que o "compartilhamento pela UIF [...] deve ser feito unicamente por meio de comunicações formais, com garantia de sigilo, certificação do destinatário e estabelecimento de instrumentos efetivos de apuração e correção de eventuais desvios"[2].

O ponto central era saber se a revelação de informações sigilosas (financeiras) pelo Conselho de Controle de Atividades Financeiras (doravante COAF) às autoridades de persecução penal por meio dos relatórios de inteligência financeira (RIFs) necessitaria ou não de autorização judicial prévia[3]. A resposta foi negativa, pois haveria permissão legal para isso (art. 15 da Lei nº 9.613/1998 [Lei de Lavagem de Dinheiro, adiante LLD]).

O tribunal também apreciou as modalidades de relatórios de inteligência financeira: espontâneos (de ofício ou de disseminação espontânea) ou "por encomenda" (disseminação a pedido), muito embora não tenha

2 Ementa do acórdão proferido pelo STF no RE 1.055.941, Tribunal Pleno, Rel. Min. Dias Toffoli, J. 04.12.2019, DJe 06.10.2020. O tema não era objeto de discussão na formulação originária do RE (cf. fl. 2699).

3 STF, RE 1.055.941, fl. 2706.

se formado maioria para que pudesse cravar uma posição sobre a admissibilidade deste último (Borges, 2021, p. 85)[4], tendo os ministros ao menos concordado quanto à "proibição da realização pelo COAF de investigações criminais prospectivas (*e.g.*, *fishing expeditions*)" (Borges, 2021, p. 90).

A discussão envolvia, essencialmente, dois diplomas legais: a LLD e a Lei Complementar nº 105/2001 (adiante, LC 105). Lembremos que, em dezembro de 2019, quando a decisão foi proferida, nem a Lei nº 13.709/2018 (Lei Geral de Proteção de Dados – LGPD) estava em vigor, nem mesmo o STF tinha julgado as ADIns 6.387-MC-Ref/DF, 6.388-MC-Ref/DF, 6.389-MC-Ref/DF, 6.390-MC-Ref/DF e 6.393-MC-Ref/DF, nas quais veio a reconhecer, em maio de 2020, que

> o tratamento e a manipulação de dados pessoais hão de observar os limites delineados pelo âmbito de proteção das cláusulas constitucionais asseguratórias da liberdade individual (art. 5º, *caput*), da privacidade e do livre desenvolvimento da personalidade (art. 5º, X e XII), sob pena de lesão a esses direitos.[5]

Tampouco a PEC 17/2019, que inclui a proteção de dados pessoais no rol dos direitos fundamentais, tinha sido aprovada.

Isso ajuda a explicar a razão pela qual, na discussão, a Corte não se fez uso, de forma geral[6], da gramática da proteção de dados pessoais, que viria a ser incorporada com mais força em seu discurso a partir de 2020. Uma outra explicação se deve, evidentemente, à exclusão, do âmbito de regência da LGPD, do tratamento de dados pessoais "realizado para fins exclusivos de [...] segurança pública" ou de "atividades de investigação e repressão de infrações penais" (art. 4º, III, da LGPD).

Fixado esse contexto, pretendo analisar, de uma forma um pouco mais ampla, a questão empregando a dogmática dos direitos fundamentais – especialmente do direito fundamental à proteção de dados pessoais –, cujos elementos centrais são a imposição de um dever de abstenção do Estado frente a direitos fundamentais e a exigência de que as intervenções sejam veiculadas por lei autorizativa proporcional.

---

4 Neste artigo, o autor faz um exame exaustivo das questões debatidas no julgamento do RE.

5 STF, ADIn 6.387-MC-Ref, Tribunal Pleno, Min. Rosa Weber, DJe 12.11.2020, J. 07.05.2020. Um exame detalhado em Souto; Rosal, 2021.

6 A exceção fica por conta do voto do Ministro Gilmar Mendes, que contém tópico dedicado ao direito fundamental à privacidade e ao sigilo de dados bancários e fiscais (fls. 3043 e ss.).

A intenção não é fazer uma crítica *ex post facto* – o que seria injusto –, mas ampliar o olhar, examinando o rendimento da aplicação dessa dogmática à matéria examinada[7]. Essa ampliação é necessária na medida em que o COAF não trata apenas dados pessoais *sigilosos*, como os financeiros, mas também (muitos) dados pessoais não protegidos por sigilo, como, por exemplo, os dados pessoais relativos a transações com joias, imóveis etc.[8] O assento constitucional dessa gramática e sua aplicação *específica* à proteção de dados na esfera penal foi desenvolvida entre nós, recentemente, por Greco (2019) e por Gleizer, Montenegro e Viana (2021), de cujas premissas partirei.

O trabalho se divide em duas grandes partes: premissas e consequências. Na primeira, serão abordadas as ferramentas normativas e dogmáticas necessárias para enfrentar a questão das comunicações feitas pelo COAF. Na segunda, essas ferramentas serão empregadas para abordar, especificamente, as comunicações do COAF para autoridades competentes.

II – AS PREMISSAS

As atividades desempenhadas pelo COAF no tratamento de dados pessoais intervêm em direitos fundamentais e têm a finalidade de convocar o aparato penal contra pessoas suspeitas da prática de infrações penais, que poderão dar ensejo a uma segunda intervenção em direitos fundamentais da mais alta gravidade. Por essas razões, o ponto de partida da análise deve ser o da teoria dos direitos fundamentais (Greco, 2019, p. 30 e ss.) ou da dogmática constitucional da proteção de dados (Gleizer; Montenegro; Viana, 2021, p. 11 e ss.; Greco; Gleizer, 2019, p. 1485-1488).

II.1 Proteção de direitos fundamentais: legalidade e proporcionalidade

II.1.a) Dever de abstenção e exigência de normas autorizativas

Direitos fundamentais são, em primeiro lugar, direitos de defesa dirigidos contra o Estado: a um direito fundamental corresponde um dever

7 Já em 1991, ou seja, há 30 anos (!), Rogall identificava o efeito que o direito constitucional e o direito de proteção de dados passaram a ter no fortalecimento dos direitos fundamentais dos afetados por investigações e ações penais (Rogall, 1991, p. 907).

8 Serão considerados dados financeiros sigilosos aqueles disciplinados na LC 105. Muitas das pessoas obrigadas a comunicar operações em espécie e suspeitas ao COAF (art. 9º da LLD) não estão abarcadas pelo espectro de proteção do sigilo financeiro estabelecido pelo art. 1º da LC 105, cujo critério de proteção não é a natureza dos dados, mas quem os trata: as instituições financeiras.

do Estado de se abster de intervir em seu âmbito de proteção[9]. Sendo a regra a abstenção e a exceção a intervenção, toda intervenção em direito fundamental tem de ser justificada. Essa justificação se assenta em três pressupostos: um formal e dois materiais. A intervenção tem de ser veiculada em lei em sentido formal, ou seja, uma autorização democrática dada pelo legislador por meio de uma norma autorizativa como exige, entre nós, o art. 5º, II, da CF[10]; "não pode atingir o núcleo dos direitos fundamentais" (Greco; Gleizer, 2019, p. 1487), e, finalmente, tem de ser proporcional, ou seja, idônea, necessária, adequada e proporcional em sentido estrito para a promoção de um fim legítimo[11].

Dentre esses pressupostos, chamo a atenção para o primeiro, pois, segundo adverte Greco, "até hoje não descobrimos todo o potencial desse dispositivo [o art. 5º, II, da CF] [...] [um] verdadeiro gigante adormecido" (2019, p. 36)[12]. A exigência de reserva legal "significa que, sem lei específica que preveja de forma relativamente clara a intervenção e lhe imponha limites materiais e procedimentais, não será lícito intervir no direito fundamental" (Greco, 2019, p. 37)[13]. Exige-se, assim, que a intervenção seja autorizada em lei em sentido formal que a veicule de forma precisa, sendo vedada sua extensão a hipóteses nela não previstas[14].

II.1.b) Normas autorizativas e normas de competência (atribuição)

E é neste ponto que a distinção entre normas *autorizativas* e normas de *competência*[15] mostra toda sua importância. Nesse sentido, quando fixa competências, "o Estado simplesmente distribui entre os seus o que incumbe a quem. Parece claro, contudo, que uma distribuição interna de tarefas

9 Greco, 2019, p. 35; Gleizer; Montenegro; Viana, 2021, p. 103, 118-123; Dimoulis; Martins, 2021, p. 179 e ss.

10 CF, art. 5º, II: "ninguém será obrigado a fazer ou deixar de fazer alguma coisa senão em virtude de lei".

11 Greco, 2019, p. 36; Gleizer; Montenegro; Viana, 2021, p. 49 e ss.; Silva, 2021, p. 103, 118-123; Dimoulis; Martins, 2021, p. 225 e ss.; Mendes; Branco, 2018, Capítulo 2, 1.3. O exame dessa exigência não será feito nesta oportunidade, mas apenas registro que não há dúvidas sobre o fato de que o Estado precisa de informações para cumprir seus deveres de prover segurança e de perseguir aqueles que praticam crimes (nesse sentido, Rogall, 1991, p. 925), o que tem um papel central no juízo de proporcionalidade.

12 No mesmo sentido, Gleizer; Montenegro; Viana, 2021, p. 41.

13 Também Silva, 2021, p. 119-120.

14 Greco; Gleizer, 2019, p. 1487. Sobre intervenções bagatelares e autorizações veiculadas por meio de cláusulas gerais, cf. Greco, 2019, p. 39-40; Gleizer; Montenegro; Viana, 2021, p. 48, 85-86, 97-99, com ulteriores referências.

15 Usarei o termo competência no decorrer do texto para ser fiel à escolha linguística feita por Greco, que não o limita à competência jurisdicional, mas abrange também normas que determinam as *tarefas*, *funções* e *atribuições* de órgãos e agentes estatais, muito embora talvez o termo atribuição fosse mais adequado ao ambiente jurídico brasileiro.

não dá a ninguém um direito de adentrar na esfera de um terceiro", o que exige, como visto, autorização clara e taxativa. Do que decorre que não se pode "derivar da existência de uma competência uma autorização" (Greco, 2019, p. 37-38).

Greco assim exemplifica essa diferença: "Incumbe ao Ministério Público proceder à investigação preliminar tão logo existam pontos de apoio fáticos no sentido do possível cometimento de um delito"; essas "normas de competência não dão ao Ministério Público, contudo, o direito de intervir na esfera jurídica de quem quer que seja", assim, se for necessária uma interceptação telefônica ou uma busca, "terá de atender aos pressupostos dos dispositivos específicos que fundamentam cada uma dessas medidas, isto é, às normas autorizativas específicas" (Greco, 2019, p. 38[16]). Ajustando o exemplo ao direito positivo brasileiro, o fato de o Ministério Público ter competência para "promover, privativamente, a ação penal pública, na forma da lei" e, para isso, "requisitar diligências investigatórias e a instauração de inquérito policial, indicados os fundamentos jurídicos de suas manifestações processuais" (art. 129, I e VIII, da CF), não autoriza seus membros a entrarem em domicílios, apreenderem objetos, interceptarem comunicações, acessarem comunicações privadas armazenadas, compartilharem dados pessoais sem autorização legal, terem acesso a dados pessoais financeiros protegidos por sigilo etc.[17] Seguindo nessa linha, o fato de "o constituinte atribuir às polícias militares a competência para exercer policiamento ostensivo e preservação da ordem pública (art. 144, § 5º, da CF) não implica, automaticamente, autorização para entrada no domicílio dos cidadãos, mesmo que isso seja necessário para o exercício de tais competências" (Gleizer; Montenegro; Viana, 2021, p. 42)[18], o mesmo valendo para as atribuições que o art. 6º do Código de Processo Penal (CPP) endereça à

16 E acrescenta: "Isso significa que, como regra, a norma tem de prever a concreta medida interventiva, e isso não apenas por meio de uma conceituação 'funcional' (obter informações, descobrir, esclarecer etc.), que é a linguagem das normas determinadoras de tarefas ou de competências, e sim com termos mais 'naturalísticos', que descrevam o concreto meio de que as instâncias de persecução se valerão para cumprir a função que lhes é legalmente atribuída" (2019, p. 39).

17 Em sentido similar, reportando-se ao § 161, I, S. 1, StPO (Código de Processo Penal alemão), que garante ao Ministério Público o poder de realizar diligências investigativas diante da suspeita da prática de crime ou de fazê-lo por meio da polícia, afirma Rogall que tal norma não autoriza a intervenção em direitos fundamentais, que só podem ser feitas pelo Ministério Público diante de normas legais de autorização veiculadas de forma clara (Rogall, 1985, p. 6-7).

18 Essa mesma dicotomia se vê no direito de polícia alemão (Greco, 2019, p. 40-41).

autoridade policial quando tiver conhecimento da prática de infração penal: apreender objetos, colher provas, ouvir pessoas etc.[19]

A força dessa verdadeira *decorrência lógica* do regime constitucional de proteção de direitos fundamentais fica clara quando pensamos que seria muito fácil atraiçoar a garantia dos direitos fundamentais do art. 5º da CF (cláusulas pétreas) se, logo à frente, o Constituinte, por meio de meras regras de competência, negasse todo o catálogo de direitos fundamentais antes consagrados. Mais concretamente: se se entendesse que normas (ainda que constitucionais) de competência consubstanciassem uma "carta branca" aos agentes estatais para que, a pretexto de "bem" cumpri-las, por exemplo, aplicassem uma sanção penal sem o devido processo legal (art. 5º, LIV, da CF) ou privassem o cidadão de parte de seu patrimônio (art. 5º, *caput* e XXII, da CF), o sistema constitucional de proteção de direitos fundamentais ruiria duas ou três páginas à frente de sua consagração, pois as regras de competência se sobreporiam às normas garantidoras de direitos fundamentais. A CF traria, já em seu âmago, uma contradição evidente, uma espécie de *back door* para a negação dos direitos que acabara de garantir. Também o disposto no art. 60, § 4º, da CF se tornaria letra morta, inútil. Daí que, se levarmos a sério o que dispõem os arts. 5º (especialmente seu inciso II[20]) e o 60, § 4º, IV, da CF, a afirmação de que *de uma regra de competência (atribuição) não decorre uma norma autorizativa* é uma decorrência lógica (também) de nosso direito (constitucional) positivo[21].

II.1.c) Proteção de dados pessoais como direito fundamental

Essas ideias se aplicarão à proteção de dados pessoais, seja ela entendida como um (novo) direito fundamental, seja como uma nova forma de proteção dos direitos gerais de personalidade[22].

Seu reconhecimento tem como marco histórico decisões tomadas pela Corte Constitucional Federal alemã23 que reconheceram que o livre

19 Manifestação clara dessa confusão no Parecer oferecido pela PGR no ARE 1.042.075, em 08.10.2021, e que trata do acesso ao conteúdo de *smartphone* apreendido pela polícia. Ali se toma a norma de competência/atribuição do art. 6º do CPP como se fosse uma norma autorizadora de intervenção em direitos fundamentais.

20 Cf. Greco, 2019, p. 40.

21 Cf. Gleizer; Montenegro; Viana, 2021, p. 41-43, com ulteriores referências, e também Greco, 2019, p. 36-37.

22 Gleizer, Montenegro e Viana (2021, p. 37-39) entendem que são vários os direitos conformando a proteção de dados pessoais; em sentido similar, Rogall, 1991, p. 926-927.

23 Sobre a origem e o desenvolvimento conceitual deste direito, além das obras já citadas, cf., ilustrativamente, entre nós, Mendes, 2020. Questionamentos acerca do acerto dessa decisão ao invocar um novo direito subjetivo em Rogall, 1985, p. 11-12, e 1991, p. 919-924.

desenvolvimento da personalidade depende do estabelecimento de limitações à obtenção, armazenamento, utilização e transferência de dados pessoais no contexto da capacidade atual de processamento, especialmente automatizado, de dados[24]. Como direito fundamental, também está sujeito a intervenções (restrições[25]), mas que devem estar previstas em lei e serem proporcionais.

É importante ter em mente que cada forma ou fase do tratamento de dados – a obtenção, o armazenamento, a utilização, a transferência etc.[26] – configura uma intervenção *autônoma* no direito à autodeterminação informacional, um direito que garante ao seu titular o controle sobre cada uso (tratamento) que é feito de seus dados. Por isso, cada forma de tratamento tem de ser objeto de autorização legal autônoma: "Uma norma que autoriza a obtenção de um dado não autoriza já automaticamente a utilização ou o armazenamento, muito menos a transferência" (Greco, 2019, p. 44[27]). Ademais, esse direito dá a seu titular o poder de saber para qual *finalidade* seus dados são coletados. Como "uma *alteração de finalidade* é um ato interventivo autônomo", que necessita de expressa autorização legal (Gleizer; Montenegro; Viana, 2021, p. 50), tanto a coleta de um dado para fins de inteligência ou segurança pública (uma finalidade) como, por exemplo, a sua transmissão e utilização para fins de persecução penal (outra finalidade) têm de estar autorizadas em lei.

II.1.d) Separação informacional

O vínculo entre *finalidade* e *autorização* de tratamento de dados mostra toda sua força por meio da ideia de *separação informacional*. No que nos interessa, da separação entre atividades de inteligência, prevenção e persecução penais: "Os dados de inteligência não podem ser usados pela polícia preventiva, os da polícia preventiva não podem ser usados repressivamente, e vice-versa, sem expressa previsão legal" (Greco, 2019, p. 45). Como a alteração de finalidade é uma nova intervenção, tem de estar prevista em lei.

24 Greco, 2019, p. 43.

25 Interessante o ponto de vista de Rogall no sentido de que esse direito não deve ser entendido como um domínio sobre dados, mas um direito limitado de disposição sobre informações pessoais, que pode variar de acordo com os riscos para os interesses pessoais (Rogall, 1985, p. 11-12).

26 Cf. LGPD, art. 5º, X.

27 Em sentido diverso, diferenciando as formas de tratamento, sustenta Rogall que a coleta (*Erhebung*) e o armazenamento (*Speicherung*) são sempre intervenções em direitos fundamentais; a utilização ou a análise (*Nutzung oder Auswertung*) de dados coletados licitamente não são intervenções, e a transmissão (*Übermittlung*) é uma intervenção, a não ser quando isso é feito dentro da mesma agência ou órgão que a coletou e desde que não haja mudança de finalidade (1991, p. 929-930).

Seria uma ousadia tentar, nesta oportunidade, delimitar os conceitos de *inteligência*, *segurança pública* e *persecução penal* em um ambiente normativo que não só não definiu (ou não o fez de forma clara) esses conceitos[28], como no qual não há uma separação clara quanto às atividades por eles abarcadas[29], e no qual há muitos resquícios normativos de nossa herança ditatorial[30]. Se a ausência de lineamentos normativos é, de um lado, perniciosa, de outro, confere certa liberdade para carregar as expressões com conteúdo semântico ligado às suas finalidades, pois finalidades diversas fundamentam amplitudes diversas de tratamento de dados, especialmente no que diz respeito à coleta.

As atividades de *inteligência* "têm como função a coleta e a análise de informações necessárias para antecipar-se a perigos ou formular políticas de segurança interna ou externa" (Gleizer; Montenegro; Viana, 2021, p. 54). Estão voltadas à precaução, razão pela qual não é necessário um ensejo (um incidente de segurança, uma prática criminosa etc.) para suas atividades. Por isso, há uma ampla margem de coleta, já que a missão aqui é, justamente, a reunião de informações, num estágio prévio à atividade policial[31]. Para que essa amplitude não se torne uma ameaça aos direitos fundamentais em um Estado Democrático[32], os órgãos de inteligência não podem agir, mas devem transmitir as informações aos órgãos de investigação/persecução penal para que estes atuem[33]. As atividades de *segurança pública* têm finalidade de proteção contra perigos, um olhar prospectivo e *preventivo* e são condicionadas por um interesse em proteger bens jurídicos contra pe-

28 Estellita; Gleizer; Montenegro, 2020.

29 O Anteprojeto de Lei de Proteção de Dados para a Segurança Pública e Persecução Penal reflete essa insegurança quando inclui serviços de inteligência tanto no conceito de segurança pública como no de persecução penal (art. 5º, XXI e XXII), no que certamente precisa de aprimoramento; todavia, a inclusão foi feita justamente em virtude da indefinição quanto a essas atividades para que não se alegasse, à frente, que os serviços de inteligência operariam ao largo de uma lei geral de proteção de dados em matéria penal.

30 Lima; Bueno; Mingardi, 2016, p. 50.

31 Greco, 2019, p. 52.

32 Muito bem percebida pelo Ministro Édson Fachin no "Caso ABIN" (STF, MC-ADIn 6.529, Tribunal Pleno, Relª Min. Cármen Lúcia, DJe 15.10.2020, p. 58).

33 A finalidade da inteligência aqui tratada é diversa daquela regulada pela Lei nº 9.883/1999, que criou o Sistema Brasileiro de Inteligência, e que, apesar de refletir essa mesma ideia de coleta e disseminação de informações, parece liminar a inteligência àquela destinada à segurança do Estado (cf. art. 1º, § 1º). Daí que Abreu, ao analisar as expressões do art. 4º, III, da LGPD, conecte as ações de inteligência justamente ao âmbito da alínea *c*, "segurança do Estado" (Abreu, 2021, p. 594). Para um apanhado histórico da regulação dos serviços de inteligência entre nós, cf. Mota; Herkenhoff; Lira, 2018. Uma crítica minuciosa à indeterminação das normas de compartilhamento de dados no âmbito da Lei nº 9.833/1999 pode ser encontrada no voto do Ministro Gilmar Mendes no "Caso ABIN" (STF, MC-ADIn 6.529, Plenário, Relª Min. Cármen Lúcia, DJe 15.10.2020, p. 88).

rigos (Gleizer; Montenegro; Viana, 2021, p. 52-53[34]). A *persecução penal*, por fim, está voltada para a confirmação de uma suspeita, tem um olhar retrospectivo e condicionado por um interesse repressivo[35].

Essas diferentes finalidades dão azo a diferentes autorizações de intervenção. As "normas autorizativas ou de faculdades autorizam ou facultam aos serviços de inteligência em primeira linha intervenções no direito à autodeterminação informacional", e, por isso, as transferências de informações de um "órgão de inteligência a um órgão de polícia é regulada por lei e limitada em vários sentidos" (Greco, 2019, p. 54-55). Na relação entre segurança pública e persecução penal, por exemplo, a polícia pode utilizar câmeras para monitorar manifestantes em uma passeata, utilizando, assim, seu efeito inibidor para evitar lesões ao patrimônio (prevenção de perigos), mas, terminada a manifestação, não pode armazenar essas imagens, a não ser que exista uma norma legal autorizando a manutenção das imagens caso algum crime tenha sido cometido e os dados sejam necessários para a persecução penal (interesse repressivo)[36].

II.1.e) Proteção de dados no Brasil

Recapitulando: o instrumental normativo até aqui reunido e que também se encontra em vigor entre nós compõe-se dos seguintes elementos: *dever de abstenção* frente a direitos fundamentais; *intervenções apenas quando autorizadas por lei proporcional*, do que decorre que *normas de competência (atribuição) não veiculam autorizações*; direito fundamental à proteção de dados pessoais/autodeterminação informacional que se insere nessa mesma gramática, exigindo *autorização em lei para cada forma de tratamento* com estrito atendimento à *finalidade legalmente prevista* para o tratamento, do que também decorre a exigência de *separação informacional*. Esse instrumental, é bom dizer, *independe* de uma lei federal de proteção de dados para o âmbito penal, muito embora uma tal lei seja desejável[37].

34 Os autores dedicam todo um capítulo ao tema das intervenções informacionais para fins de segurança pública (p. 77 e ss.) e para o qual remeto o leitor. Cf. também Greco, 2019, p. 51-52; e Abreu, 2021, p. 595.

35 Gleizer; Montenegro; Viana, 2021, p. 52.53. No mesmo sentido, Abreu, 2021, p. 595.

36 O exemplo é de Gleizer; Montenegro; Viana, 2021, p. 53-54.

37 Especialmente para regular de forma geral, expressa e detalhada alguns aspectos dessa proteção específicos desse âmbito, como, por exemplo, prazos de eliminação, exercício direto e indireto dos direitos dos titulares, autoridade nacional de proteção de dados etc. No mesmo sentido, Abreu, 2021, p. 599. Esses aspectos foram tratados pelo Anteprojeto de Lei de Proteção de Dados para a segurança pública e a persecução penal, cf. Comissão de Juristas da Câmara dos Deputados, 2020.

Foi nessa estrutura normativa constitucional de proteção de direitos fundamentais, em vigor desde 1988, que o STF assentou o reconhecimento da proteção de dados pessoais[38] e o fez reconhecendo-o como um direito fundamental no qual só se pode intervir mediante autorização prevista em lei que observe os pressupostos de proporcionalidade da intervenção (devido processo legal material, nas palavras da Corte). Isso aconteceu no julgamento das ADIns 6.387-MC-Ref/DF, 6.388-MC-Ref/DF, 6.389-MC-Ref/DF, 6.390-MC-Ref/DF e 6.393-MC-Ref/DF[39], sob relatoria da Ministra Rosa Weber, e no qual se discutia acerca da legalidade do compartilhamento de dados pessoais não anonimizados de usuários de empresas de telecomunicação com o IBGE (MP 954/2020)[40]. O ônus argumentativo que pesou sobre a Corte será bem mais leve a partir da promulgação da PEC 17/2019, aprovada pelo Senado em 20.10.2021, que inclui a proteção de dados pessoais no rol dos direitos fundamentais (art. 5º da CF).

II.2 Privacidade e proteção de dados no tratamento para as medidas de controle e prevenção da lavagem de capitais

II.2.a) Direito à privacidade, direito à proteção de dados e autodeterminação informacional

Independentemente da discussão acerca da autonomia do direito à proteção de dados[41], há consenso quanto ao fato de que esse direito vai além da ideia clássica de privacidade como direito de estar só, de se afastar da multidão - baseada numa dicotomia entre o público e o privado[42] -,

38 Rogall chama a atenção para essa mesma situação na Alemanha quando proferida a famosa "Decisão do Censo" (1983) pela Corte Constitucional, a qual, segundo ele, nada mais fez que estender ao âmbito das relações informacionais os princípios já reconhecidos no âmbito da proteção de direitos fundamentais frente a intervenções informacionais estatais (Rogall, 1991, p. 930).

39 Mas também no "Caso ABIN", STF, MC-ADIn 6.529, Tribunal Pleno, Relª Min. Cármen Lúcia, DJe 15.10.2020. Cf. especialmente o voto do Ministro Édson Fachin, p. 57, e do Ministro Gilmar Mendes, p. 77 e ss.

40 Um relato detalhado em Souto; Rosal, 2021.

41 Sobre essa discussão, cf., entre nós, o trabalho pioneiro de Doneda, 2020, *passim*. Silva registra que talvez "não haja outro direito que tenha passado por transformações tão profundas e tão rápidas em seu significado nas últimas décadas como o direito à privacidade" (Silva, 2021, p. 203). Para Bioni, 2020, p. 95, trata-se de um "novo direito da personalidade". Cf. Para Gleizer, Montenegro e Viana (2021), é uma decorrência dos direitos fundamentais que conformam a garantia ao livre desenvolvimento da personalidade (p. 38-39). A LGPD optou por assentar a proteção de dados pessoais tanto no respeito à privacidade como na autodeterminação informativa e na inviolabilidade da intimidade, da honra e da imagem (art. 2º, I, II e IV, respectivamente).

42 Bioni, 2020, p. 91-94.

para manifestar um direito de controle[43] do titular sobre os próprios dados ("pessoa-informação-circulação-controle" – Bioni, 2020, p. 94). Nesse âmbito, dados compartilhados com o público, que não gerariam por si questões atinentes à privacidade no sentido clássico, podem, "quando agregados a outros fatos (dados), revelar detalhes precisos sobre a personalidade de um indivíduo" (Bioni, 2020, p. 95). Os direitos de acesso e retificação que tem o titular de dados pessoais mesmo quando "transitam na esfera pública" (Bioni, 2020, p. 95) são um exemplo claro de que essa tutela não se confunde com a da privacidade entendida no seu sentido originário de "afastamento da multidão" (Doneda, 2020, p. 181-182)[44]. Passamos, assim, da lógica segundo a qual "dados que não são 'sigilosos'... não são protegidos" (Abreu, 2021, p. 588), para uma de proteção ampla de dados pessoais contra todas as formas de tratamento. Essas considerações são importantes no âmbito do tratamento de dados para fins de controle e prevenção da lavagem de capitais, especialmente no tratamento dado pelo COAF, objeto da decisão do STF.

II.2.b) Dados pessoais, dados pessoais sensíveis e dados pessoais sigilosos

Os dados pessoais podem ser divididos em categorias segundo certas restrições impostas a seu tratamento.

A LGPD define dado pessoal como a "informação relacionada a pessoa natural identificada ou identificável" (art. 5º, I), e dado pessoal sensível como aquele "sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, quando vinculado a uma pessoa natural" (art. 5º, II). O tratamento destes últimos está sujeito a uma série de restrições, como, por exemplo, um consentimento qualificado do titular (arts. 11 a 13 da LGPD).

No âmbito das medidas de prevenção à lavagem (adiante, PLD), há tratamento de grande quantidade de dados pessoais e de dados pessoais sensíveis. No cumprimento de suas obrigações de PLD (arts. 10, I, II, 11, I e II, da LLD), as pessoas obrigadas (art. 9º da LLD) devem proceder a várias operações de tratamento de dados pessoais em quantidades nada desprezíveis. Usando a linguagem do art. 5º, X, da LGPD, elas devem, *pelo me-*

43 Mas não de domínio ou propriedade, como sustenta Rogall, 1985, p. 11-12.

44 Também Abreu, 2021, p. 584.

nos, coletar, produzir, recepcionar, classificar, utilizar, acessar, reproduzir, transmitir, distribuir, processar, arquivar, armazenar, avaliar, controlar, modificar, comunicar, transferir e difundir dados pessoais[45].

Dentre os dados pessoais tratados pelo COAF, haverá também dados *sensíveis* e *sigilosos*.

Dados pessoais *sensíveis* serão objeto de tratamento quando relativos a "presidentes e tesoureiros nacionais, ou equivalentes, de partidos políticos", pois considerados pessoas politicamente expostas (art. 1º, § 1º, VI, da Resolução COAF nº 29, de 7 de dezembro de 2017) e relativos, portanto, à filiação a organização de caráter político (art. 5º, II, da LGPD). Nesse caso, as pessoas obrigadas devem adotar providências especiais para o acompanhamento de operações ou propostas de operações com pessoas expostas politicamente (art. 1º, *caput*, da Resolução COAF nº 29). Elas devem "dedicar especial atenção às operações ou propostas de operações envolvendo pessoa exposta politicamente, bem como com seus familiares, estreitos colaboradores e/ou pessoas jurídicas de que participem, observando, nos casos de maior risco" (art. 2º, *caput*, da Resolução COAF nº 29).

Além dessas atividades de coleta de dados pessoais feitas no âmbito do *know your customer* (KYC) ou *customer due diligence* (CDD), uma das obrigações mais relevantes para a prevenção e repressão à lavagem é a de *comunicação* de operações em espécie (COE) e de operações suspeitas (COS) ao COAF. Os dados transmitidos (para usar a linguagem da LGPD, art. 5º, X) pelas pessoas obrigadas ao cumprirem seus deveres de comunicação de operações em espécie (art. 11, II, *a*, da LLD) e suspeitas (que possam "constituir-se em sérios indícios dos crimes previstos nesta lei, ou com eles relacionar-se" – art. 11, II, *b*, da LLD), envolverá, por definição, dados pessoais. Isso é assim porque virão, necessariamente, acompanhadas de informações relacionadas às pessoas naturais envolvidas nas transações[46], pois

45 Para que se tenha uma ideia da quantidade e amplitude dos dados tratados, basta consultar a Resolução COAF nº 36, de 10 de março de 2021, que se aplica *apenas* às pessoas obrigadas sujeitas à supervisão do COAF (arts. 7º a 12). Um apanhado geral da regulação no âmbito de competência regulatória do Bacen e da CVM em Oliveira, 2020. A CVM acaba de atualizar sua regulação setorial de PLD, cf. Resolução CVM nº 50, de 31 de agosto de 2021. A amplitude dos dados tratados é objeto de preocupação no ambiente europeu há algum tempo, cf. ilustrativamente, Autoridade Europeia para Proteção de Dados, 2017; Preziosi, 2017; Malish, 2018; Khan, 2016; Peña Zafra, 2016.

46 Cf. Nota Técnica nº 40.241 do COAF referida no voto do Ministro Gilmar Mendes no RE 1.055.941, fl. 3073-3074: "O primeiro tipo de comunicação reporta operações individuais, sem a necessidade de maiores detalhamentos. A COE informa o valor da operação, a identificação do titular da conta, a pessoa que efetuou a operação, o proprietário do dinheiro e dados cadastrais bancários, tais como conta, agência, banco e cidade"; o "segundo tipo de comunicação (COS) assim se define segundo critérios emanados da lei e de

apenas elas[47] estão sujeitas à responsabilidade penal e são, portanto, os alvos finais de toda a legislação de controle, prevenção e repressão à lavagem de capitais e ao financiamento ao terrorismo[48].

Além de dados pessoais sensíveis, é da essência da atividade do COAF o tratamento de dados sujeitos a *sigilo*: um dever imposto, por lei ou ato privado, ao recipiente do dado ou informação de não revelá-lo a outras pessoas, senão àquelas expressamente autorizadas pelo titular, pelo ato privado ou pela lei[49]. O sigilo, de forma geral, restringe certas formas de tratamento que impliquem *revelação* a terceiros não autorizados do conteúdo da informação sigilosa, ou seja, na linguagem do art. 5º, X, da LGPD, a *transmissão, a distribuição, a comunicação, a transferência, a difusão*.

Há vários sigilos, médico, profissional, comercial, financeiro etc., instituídos sobre diversos fundamentos jurídicos[50], e sua violação (leia-se: a revelação da informação protegida) é conduta criminalizada, entre nós, de forma geral, pelos arts. 154 e 325 do CP[51]. Especificamente no que diz respeito ao sigilo *financeiro*, segundo opinião majoritária, serve à preservação do direito à privacidade[52], recebendo tutela penal específica, atualmente, no âmbito da LC 105[53], que incrimina a "quebra de sigilo, fora das hipóteses

regulamentos aplicáveis". O detalhamento dos dados que devem compor as duas espécies de comunicação é feito por cada regulação setorial. No caso dos bancos, por exemplo, cf. Bacen, Circular nº 3.978, de 23 de janeiro de 2020, arts. 48 a 55.

47 Exceção feita às pessoas jurídicas no âmbito dos crimes ambientais, Lei nº 9.605/1998.

48 Greenleaf; Tyree, 2017, p. 45.

49 A LGPD não define o que sejam dados sigilosos, mas indica as limitações em seu tratamento e as medidas de segurança para evitar revelação que devem ser aplicadas, cf. arts. 46 a 49.

50 Sobre os diversos fundamentos do sigilo no âmbito do § 53 do StPO (Código de Processo Penal alemão), cf. Rogall, 2018, especialmente números marginais 10 e seguintes. Sobre o sigilo médico e sua orientação à proteção de interesse individual do paciente, cf., Soares, 2020.

51 Os dois tipos penais indicam como conduta incriminada a de *revelar* segredo ou fato que deva permanecer em segredo.

52 Cf. Baltazar Júnior, 2005, p. 60; Salomão Neto, 2020, p. 678; Belloque, 2003, p. 73; Abreu, 2021, p. 584-585. Ferraz Júnior, por sua vez, parece assentar o sigilo bancário no disposto no art. 5º, XII, da CF (2020, p. 170, nota 6). No próprio RE 1.055.941, o Ministro Gilmar Mendes, por exemplo, assenta-o no art. 5º, X, da CF (fl. 3043). Como a LC 105 institui o sigilo em função das operações e dos detentores dos dados (instituições financeiras), ela não o limita às pessoas naturais, mas alcança também pessoas jurídicas. Isto poderia colocar em xeque o entendimento segundo o qual o fundamento desse sigilo é a privacidade caso se reconheça que as pessoas jurídicas não têm um direito fundamental à privacidade. Isso indicaria a correção do entendimento que vê neste sigilo uma espécie de sigilo *profissional* fundado na liberdade geral de ação, como se faz na Alemanha (Kalkbrenner; Koch, 2019, número marginal 3; que alertam para o fato de que a proteção do sigilo bancário é mais ampla quanto aos sujeitos protegidos do que a da proteção de dados pessoais, nm. 7).

53 Quanto à relação entre o art. 10 da LC 105/2001 e o art. 18 da Lei nº 7.492/1986, entende-se que os dois estão em vigor e têm abrangência distinta em função do âmbito de sujeitos abrangidos pela LC 105 e pela Lei nº 7.492/1986, cf. Baltazar, 2005, p. 172-173.

autorizadas" na lei, punindo-a com pena de reclusão, de um a quatro anos, e multa (art. 10)[54].

II.3 O tratamento de dados pessoais pelo COAF

É tempo de examinar as possíveis consequências das ideias até aqui apresentadas para o tratamento de dados pessoais pelo COAF, exame que deve ser feito em conformidade com as premissas até aqui estabelecidas e que recapitulo brevemente: (a) as competências e tarefas atribuídas ao COAF não são autorizações para intervenção em direitos fundamentais, do que decorre (b) a necessidade de que todo tratamento de dados realizado por esse órgão tenha de estar previamente autorizado por lei proporcional[55], que é aquela (c) que estabelece, de forma clara, tanto a modalidade de tratamento autorizada como a finalidade da intervenção (sendo cada nova forma de tratamento uma intervenção autônoma), e que, (d) no caso de dados pessoais protegidos por sigilo financeiro, ademais, formas de tratamento que impliquem revelação[56] (*transmissão, distribuição, comunicação, transferência, difusão*) devem cumprir rigorosamente o que dispõe a LC 105.

II.3.a) Competências e tarefas atribuídas ao COAF

A LLD criou, em 1998, o COAF com "a finalidade de disciplinar, aplicar penas administrativas, receber, examinar e identificar as ocorrências suspeitas de atividades ilícitas previstas nesta lei, sem prejuízo das competências de outros órgãos e entidades" (art. 14, *caput*)[57]. O órgão tem como atribuições, ainda, "coordenar e propor mecanismos de cooperação e de troca de informações que viabilizem ações rápidas e eficientes no combate à ocultação ou dissimulação de bens, direitos e valores" (art. 14, § 2º). Para isso, tem autorização ("poderá") para "requerer aos órgãos da Administração Pública as informações cadastrais bancárias e financeiras de pessoas

54 Outras violações graves ao regime de proteção de dados pessoais ainda carecem de tutela penal entre nós, mas já foram contempladas, por exemplo, em Portugal, na Lei nº 59/2019, arts. 53 a 60; na Itália, no Decreto Legislativo nº 51, de 18 de maio de 2018, arts. 43 a 45; na Alemanha, na *Bundesdatenschutzgesetz* – BDSG (Lei Federal de Proteção de Dados), par. 42; na Suíça, no StGB (Código Penal), arts. 143, 143-bis, 179-novies); e na Espanha, nos arts. 197 e 198 do Código Penal.

55 Em virtude dessa exigência é que o Decreto nº 9.663, de 1º de janeiro de 2019, que aprova o Estatuto do COAF, não oferece fundamento legal para tratamento de dados, razão pela qual não será objeto de exame nesta oportunidade.

56 A LC 105 fala em "conservar sigilo" (art. 1º), "violação de sigilo" (art. 1º, § 3º), "revelação de informações sigilosas" (art. 1º, § 3º, V), "quebra de sigilo" (art. 1º, § 4º), "dever de sigilo extensivo" (art. 2º), "preservação do sigilo mediante acesso restrito às partes" (art. 3º), "levantamento do sigilo" (art. 7º), "quebra de sigilo" (art. 10, que define a conduta a partir de uma metáfora).

57 Uso a redação em vigor em 20.09.2021.

envolvidas em atividades suspeitas" (§ 3º). Por fim, o órgão "comunicará às autoridades competentes para a instauração dos procedimentos cabíveis, quando concluir pela existência de crimes previstos nesta lei, de fundados indícios de sua prática, ou de qualquer outro ilícito" (art. 15).

A par dessas normas, em 7 de janeiro de 2020, foi promulgada a Lei nº 13.974, que, além de vincular o órgão ao Banco Central do Brasil (art. 2º), estabelece sua estrutura organizacional e atribuições, dentre elas as de "produzir e gerir informações de inteligência financeira para a prevenção e o combate à lavagem de dinheiro" e a de "promover a interlocução institucional com órgãos e entidades nacionais, estrangeiros e internacionais que tenham conexão com suas atividades" (art. 3º, I e II). Esse diploma legal, porém, não veicula autorizações legais para a intervenção em direitos fundamentais, mas trata da estrutura interna e funcionamento do órgão (norma de *competência*, portanto, no sentido acima empregado III, 1, *b*). Sob esse ponto de vista, por exemplo, a "produção e gestão de informações de inteligência financeira" que envolvam dados pessoais (art. 3º, I, acima) poderá ser feita desde que haja uma autorização legal, seja na LLD, seja em outra lei federal. Inexistindo autorização, a produção e a gestão não poderão envolver dados pessoais, mas, sim, por exemplo, dados anonimizados para fins estatísticos, de construção de tipologias etc. Isso vale também para a interlocução institucional com órgãos nacionais e estrangeiros (cf. *infra* IV, 3).

As tarefas atribuídas ao COAF parecem misturar elementos de *inteligência*, de *segurança pública* e de *persecução penal*. O órgão coleta e analisa informações necessárias para formular políticas de prevenção de lavagem (*inteligência*), fiscaliza o cumprimento das medidas de controle e prevenção da lavagem pelas pessoas obrigadas para, assim, prevenir perigos contra bens jurídicos (*segurança pública*) e, finalmente, se volta para o passado, ao apurar operações suspeitas de lavagem e as comunicar aos órgãos de persecução penal (*persecução penal*). Esta última faceta poderá ser ainda mais acentuada se se admitir a elaboração de RIFs a pedido (cf. *infra* IV, 2). Esse acúmulo de atribuições com finalidades diversas, que veicula uma coleta de dados de amplitude singular em nosso sistema jurídico, torna difícil alocar o órgão sob este ou aquele pilar e, consequentemente, pode gerar riscos para a proteção de dados pessoais, especialmente sob o ponto de vista da separação informacional[58].

58 O supervisor de proteção de dados europeu vem alertando para os perigos dessa falta de clareza e determinação quanto à natureza das funções das unidades de inteligência financeira desde, pelo menos, 2017: cf. European

II.3.b) Dados pessoais

Sob o ponto de vista do tratamento de dados pessoais, os arts. 11, 14 e 15 da LLD sugerem o seguinte regime: a) o art. 11 veicula a norma autorizativa que autoriza as pessoas obrigadas (do art. 9º) a transmitirem ao COAF dados pessoais não sujeitos a sigilo financeiro;[59] b) o art. 14, *caput*, lido com benevolência (já que a linguagem se aproxima mais da de uma norma de competência do que de uma norma de autorização[60]), autoriza o COAF ao tratamento ("receber, examinar e identificar") interno de dados pessoais recebidos das pessoas obrigadas exclusivamente para a finalidade de apuração da ocorrência de suspeita de prática de lavagem de dinheiro e financiamento ao terrorismo ("as ocorrências suspeitas de atividades ilícitas previstas nesta lei"), vedado o tratamento para qualquer outra finalidade;[61] c) uma especial forma de tratamento, a comunicação (também chamada de "disseminação"), só pode ser realizada *pelo* COAF *para* as "autoridades competentes para a instauração dos procedimentos cabíveis, quando concluir pela existência de crimes previstos nesta lei, de fundados indícios de sua prática, ou de qualquer outro ilícito" (art. 15); d) o dever imposto ao órgão de "coordenar e propor mecanismos de cooperação e de troca de informações que viabilizem ações rápidas e eficientes no combate à ocultação ou dissimulação de bens, direitos e valores" (art. 14, § 2º)[62], não parece veicular norma autorizativa para compartilhamento de dados pessoais (sigilosos ou não) com pessoas que não as indicadas no art. 15, pois o que a lei autoriza é que o órgão *coordene* e *proponha*, mas não que *troque* informações, ou seja, que as transmita, distribua, comunique, transfira, difunda ou dissemine.

Data Protection Supervisor, 2017, parágrafo 52; e, mais recentemente, European Data Protection Supervisor, 2020, parágrafos 35 e 36.

59 "Art. 11. As pessoas referidas no art. 9º: [...] II – deverão comunicar ao COAF, abstendo-se de dar ciência de tal ato a qualquer pessoa, inclusive àquela à qual se refira a informação, no prazo de 24 (vinte e quatro) horas, a proposta ou realização: a) de todas as transações referidas no inciso II do art. 10, acompanhadas da identificação de que trata o inciso I do mencionado artigo; e b) das operações referidas no inciso I."

60 Neste sentido, cf. o § 28 da *Geldwäschegesetz* (adiante, GwG), que detalha, em 13 incisos, todas as atividades da UIF alemã, seguido do § 29, que disciplina o tratamento de dados pessoais pela UIF.

61 "Art. 14. Fica criado, no âmbito do Ministério da Economia, o Conselho de Controle de Atividades Financeiras – COAF, com a finalidade de disciplinar, aplicar penas administrativas, receber, examinar e identificar as ocorrências suspeitas de atividades ilícitas previstas nesta Lei, sem prejuízo das competências de outros órgãos e entidades." (Redação dada pela Medida Provisória nº 886, de 2019)

62 Art. 15: "§ 2º O COAF deverá, ainda, coordenar e propor mecanismos de cooperação e de troca de informações que viabilizem ações rápidas e eficientes no combate à ocultação ou dissimulação de bens, direitos e valores".

II.3.c) Dados pessoais financeiros sigilosos

Além de dados pessoais comuns e sensíveis, o COAF recebe e trata também dados pessoais financeiros *sigilosos*, submetidos, até 2001, ao regime do art. 38 da Lei nº 4.595/1964, e, a partir de 2001, ao regime da LC 105. A regra é que as operações ativas e passivas e os serviços prestados pelas instituições financeiras indicadas no art. 1º, § 1º, da LC 105 sejam sigilosos, no sentido de que as informações a elas relativas *não possam ser reveladas* (transmitidas, distribuídas, comunicadas, transferidas, difundidas ou disseminadas) a terceiros.

A relação entre a LLD (e suas alterações) e a LC 105 habita o problemático ambiente das relações entre leis ordinárias e leis complementares. A partir de 1988, o art. 192, *caput*, da CF passou a determinar que o sistema financeiro nacional fosse regulado por lei complementar. Por essa razão, a Lei nº 4.595, de 31 de dezembro de 1964, que tratava da matéria, foi recepcionada pela nova ordem constitucional como lei complementar[63]. E era seu art. 38 que, até 2001, instituía o sigilo financeiro. Em 2001, como dito, sobreveio nova lei complementar disciplinando, de forma mais detalhada, o sigilo financeiro, a LC 105.

Muito embora o STF tenha reconhecido, por maioria, que não existe, em regra, hierarquia entre uma lei ordinária e uma lei complementar, há certa distribuição constitucional material (*ratione materiae*) entre as espécies legislativas[64]: uma reserva constitucional de lei complementar limitada a certas matérias. Assim, uma lei ordinária poderia dispor em sentido contrário do disposto em uma lei complementar e, assim, revogar seus dispositivos, *desde que* não trate de matéria privativa desta[65]. *A contrario sensu*, uma lei complementar que trate de matéria a ela não reservada pela CF vale como lei ordinária[66].

E por que isso importa?

Porque, tendo sido o sigilo financeiro estabelecido em lei complementar, se houver, para essa matéria, reserva constitucional de lei comple-

63 Baltazar, 2005, p. 73.

64 STF, RE 377.457, Tribunal Pleno, Rel. Min. Gilmar Mendes, DJe 18.11.2008.

65 STF, RE 377.457, Tribunal Pleno, Rel. Min. Gilmar Mendes, DJe 18.11.2008, fl. 1819.

66 STF, RE 377.457, Tribunal Pleno, Rel. Min. Gilmar Mendes, DJe 18.11.2008, fl. 1834. A complexidade dessa discussão se mostra com toda sua força na discussão travada no plenário por ocasião deste julgamento e que não pode ser captada nesta oportunidade.

mentar[67], as normas da LLD não revogariam nem as da Lei nº 4.595/1964 (até 2001), nem as da LC 105 (após 2001), que regeriam, soberanas, as autorizações para a revelação de dados financeiros sigilosos. Em outras palavras: a revelação de dados financeiros sigilosos *para* o COAF e *pelo* COAF dependeriam das autorizações expressas da LC 105. Nesse caso, tanto as comunicações de operações (em espécie e suspeitas) feitas pelas pessoas obrigadas *para* o COAF contendo dados financeiros sigilosos como *do* COAF para as autoridades competentes para a persecução penal (por meio de RIFs no sentido do art. 15 da LLD) dependeriam de norma autorizativa na LC 105.

Sob esse entendimento, teríamos o seguinte quadro à luz da LC 105: a) haveria uma autorização para "a comunicação, às autoridades competentes, da prática de ilícitos penais ou administrativos, abrangendo o fornecimento de informações sobre operações que envolvam recursos provenientes de qualquer prática criminosa" (art. 1º, § 3º, IV, da LC 105), aplicável a quaisquer pessoas submetidas à LC 105, dentre elas parte das pessoas obrigadas do art. 9º da LLD. Essa seria a norma que veicularia autorização legal para a comunicação de operações suspeitas (COS) pelas pessoas obrigadas sujeitas à LC 105 ao COAF; b) haveria autorização para que o Bacen e a CVM, no exercício de suas atribuições e quando verificassem "a ocorrência de crime definido em lei como de ação pública, ou indícios da prática de tais crimes", informassem o Ministério Público sobre tais fatos juntando à comunicação os documentos necessários à apuração ou comprovação dos fatos (art. 9º, *caput*, da LC 105); ou seja, haveria uma autorização para comunicação não ao COAF, mas ao MP, sobre eventual suspeita de prática de crime; c) haveria autorização para que o Bacen e a CVM e demais órgãos de fiscalização (mas não as instituições financeiras), nas áreas de suas atribuições, fornecessem ao COAF "as informações cadastrais e de movimento de valores relativos às operações previstas no inciso I do art. 11 da referida lei", que são as que "possam constituir-se em sérios indícios dos crimes previstos nesta lei, com eles relacionar-se" (art. 11, I, da LLD). O dispositivo, portanto, só alcança operações suspeitas (art. 2º, § 6º, da LC 105); d) quanto às comunicações de operações em espécie, disciplinadas em dois dispositivos da LLD[68], a LC 105 não teria veiculado autorização para que as pes-

67 Um tema que merece estudo mais aprofundado, inviável aqui.

68 O art. 10, II, e o art. 11, II, que distingue, em suas alíneas *a* e *b*, entre operações em espécie (que geram as comunicações denominadas COE) e as suspeitas (que geram comunicações denominadas COS).

soas obrigadas pela LLD, mas submetidas ao sigilo financeiro, fizessem tais comunicações, a não ser que se entenda que toda operação em espécie é suspeita, em um sentido mais amplo deste termo, diverso do sentido que lhe é dado pela LLD; e) por fim, a LC 105 instituiu o sigilo de dados financeiros, mas não tratou da comunicação do COAF *para* as "autoridades competentes" (art. 15 da LLD) que contenham dados protegidos por sigilo financeiro, como fez, expressamente, quanto ao Bacen, à CVM e à Receita Federal[69]. Esse pernicioso silêncio pode dar azo a disputas quanto à possibilidade de o COAF incluir dados cobertos por sigilo financeiro em seus relatórios para autoridades de persecução penal.[70]

A questão ficou ainda mais complexa com o advento da mudança do COAF para âmbito do Bacen por força da Lei nº 13.974, de 7 de janeiro de 2020. Se se entender que, a partir de então, o órgão e seus agentes fazem parte do Bacen, a eles se estende o dever de sigilo (art. 2º, § 5º, da LC 105) e também a autorização para a comunicação prevista no art. 9º, o que tornaria inquestionável a possibilidade de que os RIFs contivessem dados protegidos por sigilo financeiro. A complexidade criada pelo art. 2º da Lei nº 13.974, que "vinculou administrativamente" o COAF ao Bacen, mas manteve sua "autonomia técnica e operacional", está a merecer melhor exame, especialmente sob o viés da separação informacional.

Se, porém, se entender que a matéria do sigilo financeiro, apesar de veiculada em lei complementar, não está a ela reservada, então tanto as normas da Lei nº 4.595/1964 como as da sua sucessora, a LC 105, valeriam como lei ordinária, e as normas da LLD teriam o condão de regular a matéria, criando permissões de revelação desses dados não contempladas na LC 105. Neste caso, o art. 11 da LLD, na redação recebida por força da Lei nº 12.683/2012, seria a base legal que autorizaria todas as pessoas obrigadas – dentre elas também aquelas sujeitas à LC 105, que é anterior à Lei nº 12.683/2012 – a comunicar dados pessoais financeiros sigilosos ao COAF; e o art. 15 da LLD autorizaria este órgão a comunicá-los, nos RIFs, às autoridades competentes.

Até aqui, a bem ver, tratamos de revelação de dados pessoais *financeiros sigilosos* limitados a certas operações (suspeitas, em espécie, indica-

69 Cf. LC 105/2001, art. 1º, § 3º, VI, c/c, especialmente, com o art. 9º.

70 Essa incerteza não é um privilégio do Brasil, como registra Maillart em tom crítico, quanto aos cinco países analisados em abrangente estudo comparativo sobre a regulação da lavagem de capitais concluído em 2020 (Vogel; Maillart, 2020, p. 848).

tivas da prática de crime), do que decorre uma delimitação clara dos dados que podem ser revelados em termos de tempo, espaço, pessoas envolvidas, montantes, espécie de operação ou serviços, pois a *finalidade* é clara. Embora, como dito, configurem intervenções no direito à proteção de dados pessoais e na privacidade, a severidade dessa intervenções, em si mesmas, é mais limitada do que a revelação de dados financeiros não conectados a certas operações, mas, sim, a lapsos temporais. Neste último caso, trata-se de revelar aspectos amplos não só da privacidade da pessoa afetada, mas de sua intimidade, além de dados de terceiros insuspeitos[71]. Enquanto a comunicação de uma prática criminosa está limitada (finalidade) apenas às informações necessárias para a comprovação da suspeita, por exemplo, o recebimento de certa quantia, em certa data, envolvendo certas pessoas etc.; o que conhecemos por "quebra" do sigilo bancário por ordem judicial ou de comissões parlamentares de inquérito implica revelação indiscriminada da privacidade e da intimidade tanto da pessoa que é alvo da "quebra" como de terceiros insuspeitos dentro do espectro temporal coberto pela determinação. Trata-se, assim, de uma intervenção muito mais severa no âmbito do direito fundamental à privacidade e é por essa razão que, nestes casos, a LC 105 não veicula uma autorização direta para que agentes públicos os acessem, mas exige uma autorização judicial individualizada (*reserva de jurisdição* – art. 1º, § 4º) ou a aprovação por órgão colegiado no âmbito das comissões parlamentares de inquérito (art. 4º, § 2º), e, ademais, determina que o caráter sigiloso seja mantido mediante "acesso restrito às partes, que delas não poderão servir-se para fins estranhos à lide" (art. 3º, *caput*, da LC 105)[72].

III – CONSEQUÊNCIAS

Estabelecidas todas essas premissas, cumpre delas extrair as consequências para aquela que é a questão central examinada neste texto: os limites

71 Cf. Estellita; Gleizer, 2020.

72 Este é só um dentre os inumeráveis fundamentos que evidenciam a ilegalidade da determinação feita pela PGR, em meados de 2020, às Forças-Tarefa da Operação Lava-Jato no Rio de Janeiro, São Paulo e Curitiba para que lhe entregassem "todas as bases da dados estruturados e não estruturados utilizadas e obtidas em suas investigações, por meio de sua remessa atual, e para dados pretéritos e futuros, à Secretaria de Perícia, Pesquisa e Análise do gabinete do procurador-geral da República" (Disponível em: https://g1.globo.com/politica/noticia/2020/08/03/fachin-revoga-decisao-de-toffoli-que-permitia-compartilhamento-de-dados-entre-pgr-e-forcas-tarefa-lava-jato.ghtml. Acesso em: 11 set. 2021), seguida da Portaria Conjunta PGR/MPF – CMPF nº 1, de 7 de janeiro de 2021. Outras considerações em Luz, 2021. A determinação é objeto de disputa no STF, na RCL 42.050, que tramita sob sigilo.

da transmissão, distribuição, comunicação, transferência e difusão (ou disseminação, no jargão do COAF) de dados pessoais *pelo* COAF.

III.1 RIFs de ofício ("disseminação espontânea")

Como visto, tanto a LLD como a LC 105 contêm autorizações[73] para que o COAF receba, classifique, utilize, processe, arquive, armazene, avalie e controle dados pessoais, sigilosos ou não. Quanto às operações de transmissão, distribuição, comunicação, transferência e difusão (que evidentemente incluem a "disseminação"), norma autorizativa para essas formas de tratamento é o art. 15 da LLD, que as limita à comunicação às autoridades competentes, quando o próprio órgão concluir pela existência ou fundados indícios da prática de crimes previstos na LLD, ou qualquer outro ilícito[74]. Nada mais é dito sobre a forma e o conteúdo dessas comunicações[75].

III.2 RIFs a pedido ("disseminação a pedido")

Os dois diplomas legais que regulam a atividade do COAF não parecem lhe impor um dever de compartilhar dados pessoais *a pedido de autoridades públicas*[76].

Com relação a eventuais pedidos de representantes do Ministério Público, muito embora seja sua "função institucional" "promover, privativamente, a ação penal pública, na forma da lei" e "requisitar diligências investigatórias e a instauração de inquérito policial, indicados os fundamentos jurídicos de suas manifestações processuais" (art. 129, I e VIII, da CF)[77], dessas atribuições não derivam autorizações para intervenção em direitos

73 Muito embora com linguagem fraca em termos da clareza e da determinação que se deve exigir das normas autorizativas de intervenções em direitos fundamentais (cf. acima II, 1, *a* e *b*).

74 A referência a "qualquer outro ilícito", prevista no final do art. 15, *caput*, da LLD, merece reparo, pois poderia levar a transformar o órgão de combate e prevenção à lavagem de dinheiro e ao financiamento do terrorismo em agência de inteligência de quaisquer ilícitos, penais ou administrativos, praticados no País. Sobre a importância de rigorosa observância da finalidade do tratamento de dados pelas unidades de inteligência financeira, cf. European Data Protection Supervisor, 2020, parágrafos 8 a 10.

75 O dispositivo é lacônico, beirando o descumprimento do princípio da clareza e determinação e merece atenção do legislador. Nesse sentido, vale conferir o detalhado § 32, 5, GwG.

76 Em sentido similar, Bottini, 2021. Sobre a situação nos cinco países objeto da pesquisa coordenada por Vogel e Maillart, cf. Vogel; Maillart, 2020, p. 848-849.

77 A Lei Complementar nº 75/1993 repete essa regra de competência (art. 6º, V) e contém outra regra de competência no art. 7º. A Lei Orgânica Nacional nº 8.625/1993 reproduz essa linguagem no art. 26, II e IV.

fundamentais, as quais, como dito e repetido, têm de ser autorizadas por lei (art. 5º, II, da CF)[78].

O Ministério Público tem, assim, competência para fazer essas requisições, mas o atendimento que implique tratamento de dados pessoais só pode ser efetuado se houver autorização legal proporcional. Um entendimento que autorizasse aos membros do MP a obtenção direta, junto às instituições financeiras ou ao COAF, de informações cobertas por sigilo financeiro permitiria que, como dito, pela porta dos fundos (*back door*), fosse corroído o regime constitucional (e infraconstitucional) de proteção de direitos fundamentais. Isso implicaria, ademais, verdadeira *fusão informacional* entre os dois órgãos[79], pois, por essa via, o Ministério Público obteria acesso a um imenso conjunto de dados que o legislador outorgou *apenas* ao COAF[80]. Esses mesmos limites valem, logicamente, para as autoridades policiais.

Conforme explicações fornecidas pelo COAF ao STF no RE aqui analisado, o Sistema SEI-C, além de ser utilizado para as comunicações do COAF para as autoridades competentes, também é utilizado como um canal de comunicação na via oposta: das autoridades competentes para o COAF. Nele, essas autoridades registram dados sobre pessoas investigadas, os crimes dos quais são suspeitas e a descrição do modo como os teriam praticado. Essas novas informações, incorporadas ao acervo informacional que o COAF já tem[81], podem ser objeto de análise, com a conclusão da existência de operações suspeitas de lavagem. Neste caso, o COAF pode, naturalmente, gerar um RIF e comunicar as operações suspeitas individualizadas às autoridades competentes, nos exatos termos da autorização que lhe é dada pelo art. 15 da LLD. Esta foi, inclusive, a posição do relator do RE 1.055.941 (cf. fl. 2738 do acórdão) no STF e que parece acertada, pois aqui, na verdade, não se

78 A teoria dos poderes implícitos discutida no STF quando chamado a decidir se membros do Ministério Público poderiam investigar não é incompatível com o que aqui se afirma, pois seus poderes de investigar só podem ser exercidos desde que respeitados os direitos e garantias fundamentais (cf. STF, RE 593.727, Tribunal Pleno, Rel. Min. Cezar Peluso, DJ 08.09.2015).

79 Em sentido similar, Greco; Leite, 2019.

80 Badaró chama a atenção para esse ponto considerando equivocada a decisão do STF no RE por não separar adequadamente "quem detém a informação" de "quem detém o poder de persecução penal" (Badaró, 2021). Em sentido diverso, aparentemente, Bottini, 2021.

81 Questão que está a merecer atenção diz respeito à necessidade de maior transparência quanto às fontes de dados. Sobre isso, cf. Bialski; Vento; Messina; Wiegerinck, 2021.

trata de um RIF ou "disseminação" "a pedido", mas de RIF de ofício ("disseminação espontânea")[82].

III.3 Intercâmbio internacional entre UIFs

A questão é bem mais delicada quando se trata do intercâmbio de dados pessoais (dentre eles, também financeiros sigilosos) com unidades de inteligência financeira de outros países, como no âmbito do chamado Grupo Egmont. Segundo informações públicas do próprio COAF[83], a cooperação com as UIFs estrangeiras dá-se no âmbito do próprio SEI-C, onde podem ser transmitidos e transferidos (e também obtidos) dados como comunicações de operações suspeitas, comunicações de operações em espécie, identificação de pessoas naturais e jurídicas (suspeitas e relacionadas a suspeitas), identificação de beneficiários finais, de sócios e representantes de empresas, participações societárias, lista de bens móveis e imóveis, histórico criminal etc.

Como dito, nem o dever imposto ao órgão de "coordenar e propor mecanismos de cooperação e de troca de informações" (art. 14, § 2º), nem mesmo sua competência para, "em todo o território nacional", "promover a interlocução institucional com órgãos e entidades nacionais, estrangeiros e internacionais que tenham conexão com suas atividades" (art. 3º, II, da Lei nº 13.974/2020) lhe franqueiam autorização legal para compartilhar dados pessoais (sigilosos ou não) com pessoas que não as do art. 15 da LLD. A cooperação e a troca de informações, assim, não poderia envolver dados pessoais (sigilosos ou não), a não ser que houvesse diploma legal (reserva de lei) autorizando o órgão a fazê-lo[84], como poderia ser, por exemplo, um

82 Razão pela qual seria recomendável não o denominar de "RIF decorrente de intercâmbio", expressão que sugere um RIF produzido por provocação de autoridades públicas, tratamento de dados que, como visto, não está autorizada pela LLD.

83 COAF – CONSELHO CONTROLE DE ATIVIDADES FINANCEIRAS. Inteligência Financeira: Aspectos práticos do intercâmbio internacional via Rede Egmont. [s.l.: s.n., s.d.]. Disponível em: https://www.youtube.com/watch?v=i5N_LqLmewI. Acesso em: 8 ago. 2021.

84 Segundo Teixeira, Wehrs e Madruga, a troca de dados entre unidades de inteligência financeira integrantes do Grupo Egmont não é regulada por convenção ou tratado internacional, mas apenas pelas normas e princípios do próprio Grupo (Teixeira; Wehrs; Madruga, 2019, p. 22). O *site* do órgão igualmente não indica uma base legal para a cooperação, cf. https://www.gov.br/coaf/pt-br/acesso-a-informacao/Institucional/articulacao-institucional/ariticulacao-internacional-em-pld-ft. Acesso em: 11 set. 2021. As Convenções de Mérida e de Palermo incentivam a cooperação, mas sempre observada a legislação interna (Convenção de Mérida, art. 14, 1, *b*; Convenção de Palermo, art. 7, 1, *b*).

tratado internacional incorporado ao nosso direito positivo com hierarquia de lei federal[85].

Poder-se-ia, talvez, deduzir essa permissão do próprio art. 15, que não teria limitado o rol de autoridades competentes às domésticas, a par de estar em conformidade com as recomendações do GAFI. Esse entendimento, porém, é questionável por algumas razões.

De um lado, o GAFI é um órgão intergovernamental, um fórum, que emite orientações e diretrizes para o combate à lavagem de capitais e financiamento do terrorismo. Suas recomendações não têm impacto vinculante em nosso direito positivo ("*soft law*"). De outro, dentro do regime de proteção de direitos, também as transferências internacionais devem ser veiculadas por norma legal proporcional, pois se trata de intervenção especialmente severa envolvendo alvos que estarão privados da tutela jurisdicional prestada pelo Poder Judiciário brasileiro[86], feita sem exigência de que o receptor tenha um nível adequado de proteção de dados pessoais[87]. Além disso, "autoridades competentes para a instauração dos procedimentos cabíveis" têm sido compreendidas como autoridades competentes para a investigação e persecução penal de infrações penais[88], algo que nem todas UIFs são, como serve de exemplo o próprio COAF. Quanto aos dados pessoais financeiros sigilosos, a LC 105, além de dever ser interpretada de forma restritiva[89], não autorizou o COAF a transmitir esses dados para autoridades estrangeiras. Por fim, uma eventual invocação de atos normativos inferiores à lei (decretos, convênios, acordos de cooperação etc.) não supriria o que determina o art. 5º, II, da CF, pois "não há autorização constitucional a agentes não

85 Pressuposto ligado à validade da cooperação ativa, mas que não necessariamente habilitaria o Brasil a receber dados de Estados estrangeiros, dada a inexistência de infraestrutura legal de proteção de dados pessoais nesse âmbito. Nesse sentido, cf. Aras, 2020, p. 14-31, p. 26 e ss. Um panorama do direito positivo brasileiro, do ambiente legal da União Europeia e da Convenção de Budapeste em Domingos, Abreu e Silva; Oliveira, 2020, p. 140-162. Sobre as exigências de adequação ao *standard* de proteção europeu, cf. Morán Martínez, 2020, p. 163-196 e European Data Protection Board, 2021.

86 No mesmo sentido, Teixeira; Wehrs; Madruga, 2019, p. 24. Os autores lembram, ilustrativamente, as autorizações expressas e detalhadas dadas pela lei alemã de lavagem de dinheiro, bem como pela portuguesa.

87 Como é feito, por exemplo, no âmbito da União Europeia, onde se exige uma decisão de adequação prévia da Comissão Europeia (art. 36 da Diretiva UE nº 680/2018). Cf. European Data Protection Board, 2021. Ainda que se leve em conta o disposto no art. 33 da LGPD, o inciso VII, que poderia amparar a transferência, não faz tal exigência ao destinatário dos dados. De outro lado, o inciso III não se aplica à hipótese ora analisada, pois o Grupo Egmont não é um órgão público e nem sequer há um acordo de cooperação internacional que permita a invocação do inciso VI.

88 Cf. RE 1.055.941, voto do Ministro Gilmar Mendes, fl. 3073, 3075.

89 Cf. STF, RE 1.055.941, voto do Ministro Gilmar Mendes, fl. 3049, 3072.

parlamentares para que decidam, à margem do processo democrático, impor restrições a direitos de defesa que valem, principalmente, contra eles"[90].

Em um mundo no qual já se reconhece que não há dados irrelevantes, no qual a proteção de dados pessoais é um direito fundamental, no qual os diplomas legais nacionais e regionais que regulam a proteção de dados impõem uma série de requisitos para o intercâmbio internacional de dados e criam diversos mecanismos de proteção dos titulares afetados[91], uma tal interpretação do art. 15 merece ser revista, por obsoleta e incompatível com essa nova realidade[92].

Essas objeções não impedem, evidentemente, que uma autorização legal expressa e proporcional venha a ser veiculada pelo Poder Legislativo. Até lá, porém, a prática, nos limites das parcas informações acessíveis publicamente, não parece ser admissível.

IV – À GUISA DE CONCLUSÃO

Ao fim deste exercício, fica patente que há muito a fazer em termos de observância do princípio da legalidade e do respeito à proteção de dados pessoais num ambiente cujo tratamento de dados é tão amplo e intenso como o do COAF e feito para fins que podem conduzir à privação da liberdade de seus alvos.

Para que esse labor possa ser feito sem causar prejuízos irreparáveis às atividades de inteligência, de segurança pública e de persecução penal, poderíamos muito bem nos aproveitar daquilo que os alemães chamam de bônus de transição[93], e que nós chamamos de modulação de efeitos[94], con-

90 Gleizer; Montenegro; Viana, 2021, p. 44. Como dito, o Decreto nº 9.663, de 1º de janeiro de 2019, que aprova o Estatuto do COAF, não oferece fundamento legal para a transmissão, a distribuição, a comunicação, a transferência, a difusão ou a disseminação de dados pessoais, sigilosos ou não, devendo a troca limitar-se a tipologias, dados estatísticos anonimizados etc. (cf. art. 16).

91 Apenas ilustrativamente, cf. LGPD, arts. 33 a 36; Regulamento (UE) nº 2.016/679 do Parlamento Europeu e do Conselho de 27 de abril de 2016 (Regulamento Geral sobre a Proteção de Dados), arts. 41 a 45; Diretiva (UE) nº 2.016/680 do Parlamento Europeu e do Conselho de 27 de abril de 2016, arts. 35 a 40. O já mencionado Anteprojeto também sugeriu disciplina para a matéria em seus arts. 53 a 58.

92 Incompatibilidade que, a rigor, afeta, como um todo, o ambiente brasileiro relativo à prevenção, investigação, detecção ou repressão de infrações penais ou execução de sanções penais. O Brasil, neste momento, não reúne condições mínimas para conseguir uma decisão de adequação para ser recipiente de dados oriundos do ambiente europeu (cf. o guia para adequação preparado pela autoridade europeia de proteção de dados, European Data Protection Board, 2021).

93 Trata-se da concessão de um prazo ao legislador para que implemente as exigências relativas à proteção de dados. Cf. Greco; Leite, 2019; Greco, 2019, p. 47; Wolter, 2019, p. 167.

94 O Anteprojeto contém uma cláusula temporal de adequação no prazo da *vacatio* ali sugerida, que é de 365 dias (arts. 67 e 68). Talvez esse prazo deva ser ampliado dado o trabalho que deverá ser feito por juristas e parlamentares para adequar a legislação às exigências constitucionais no trato com dados pessoais.

cedendo um prazo razoável para a adaptação da legislação às exigências da proteção de dados pessoais.

Enquanto essa legislação não vem e no atual ambiente, no qual apenas se aguarda a promulgação da PEC 17/2019, no qual a LGPD já está em vigor e, por fim, no qual o STF já tomou importantes decisões em prol da proteção de dados pessoais, é de se esperar que a Corte adote uma abordagem mais rigorosa tanto no que diz respeito à exigência de autorização legal como no que tange à estrita observância dos limites legais para o tratamento de dados pessoais nos âmbitos da inteligência, da segurança pública e da persecução penal. No que interessa a este artigo, isso poderá ter impactos na questão da admissibilidade dos RIFs a pedido e na do intercâmbio internacional entre UIFs, podendo também conduzir a um reexame dos limites do compartilhamento de dados financeiros sigilosos.

REFERÊNCIAS

ABREU, Jacqueline de Souza. Tratamento de dados pessoais para segurança pública: contornos do regime jurídico pós-LGPD. In: MENDES, Laura Schertel; DONEDA, Danilo; SARLET, Ingo Wolfgang *et al.* (Org.). *Tratado de proteção de dados pessoais*. São Paulo: Forense, p. 583-603, 2021.

ARAS, Vladimir Barros. A título de introdução: segurança pública e investigações criminais na era da proteção de dados. In: DE MENDONÇA, Andrey Borges; CAPANEMA, Walter Aranha; ARAS, VLADIMIR BARROS, Augusto *et al.* (Org.). *Proteção de dados pessoais e investigação criminal*. Brasília: ANPR, p. 14-31, 2020.

AUTORIDADE EUROPEIA PARA PROTEÇÃO DE DADOS. Síntese do Parecer da Autoridade Europeia para a Proteção de Dados relativo à proposta da Comissão que altera a Diretiva (UE) nº 2015/849 e a Diretiva nº 2009/101/CE – Acesso a informações sobre os beneficiários efetivos e implicações para a proteção de dados, 2017.

BADARÓ, Gustavo. O debate constitucional sobre privacidade, intimidade e proteção de dados no Brasil. In: CRUZ, Francisco Brito; SIMÃO, Bárbara (Ed.). *Direitos fundamentais e processo penal na era digital* [livro eletrônico]. São Paulo: InternetLab, 2021 (sem paginação).

BALTAZAR JUNIOR, José Paulo. *Sigilo bancário e privacidade*. Porto Alegre: Livraria do Advogado, 2005.

BELLOQUE, Juliana Garcia. *Sigilo bancário*: análise crítica da LC 105/2001. São Paulo: Revista dos Tribunais, 2003.

BIALSKI, André; VENTO, Antonio; MESSINA, Eduardo; WIEGERINCK, Oliver. *Transparência no tratamento de dados por UIFs*: em busca de um *benchmark*. São Paulo: FGV-Data Privacy Brasil, 2021 (no prelo).

BIONI, Bruno Ricardo. *Proteção de dados pessoais*: a função e os limites do consentimento. 2. ed. Rio de Janeiro: Forense, 2020.

BORGES, Ademar. O relatório de inteligência financeira como meio de obtenção de prova no processo penal. *Revista Brasileira de Ciências Criminais*, São Paulo, v. 176, p. 69-105, 2021.

BOTTINI, Pierpaolo Cruz. Os limites da atuação do COAF. *Consultor Jurídico*, 29.03.2021. Disponível em: https://www.conjur.com.br/2021-mar-29/direito-defesa-limites-atuacao-coaf. Acesso em: 8 ago. 2021.

COAF – CONSELHO CONTROLE DE ATIVIDADES FINANCEIRAS. *Inteligência financeira*: aspectos práticos do intercâmbio internacional via Rede Egmont. [s.l.: s.n., s.d.]. Disponível em: https://www.youtube.com/watch?v=i5N_LqLmewI. Acesso em: 8 ago. 2021.

COMISSÃO DE JURISTAS DA CÂMARA DOS DEPUTADOS. Anteprojeto de Lei de Proteção de Dados para segurança pública e persecução penal, 2020.

DIMOULIS, Dimitri; MARTINS, Leonardo. *Teoria geral dos direitos fundamentais*. 8. ed. São Paulo: Revista dos Tribunais, 2021.

DOMINGOS, Fernanda Teixeira Souza; ABREU E SILVA, Melissa Garcia Blagitz; OLIVEIRA, Neide M. Cavalcanti Cardoso de. Transferência internacional de dados pessoais para fins de investigações criminais à luz das leis de proteção de dados pessoais. In: ARAS, Vladimir Barros; DE MENDONÇA, Andrey Borges; CAPANEMA, Walter Aranha (Org.). *Proteção de dados pessoais e investigação criminal*. Brasília: ANPR, p. 140-162, 2020.

DONEDA, Danilo. *Da privacidade à proteção de dados*. 2. ed. São Paulo: Revista dos Tribunais, 2020.

ESTELLITA, Heloisa; GLEIZER, Orlandino. A investigação penal de insuspeitos. *Folha de S. Paulo*, p. A3, 2020.

_____; _____; MONTENEGRO, Lucas. Por um direito de segurança pública. *Estadão*, 05.10.2020. Disponível em: https://politica.estadao.com.br/blogs/fausto-macedo/por-um-direito-de-seguranca-publica/. Acesso em: 9 set. 2021.

EUROPEAN DATA PROTECTION SUPERVISOR. *Opinion 1/2017* – EDPS Opinion on a Commission Proposal amending Directive (EU) 2015/849 and Directive 2009/101/EC Access to beneficial ownership information and data protection implications. [s.l.: s.n.], 2017.

_____. Opinion 5/2020 on the European Commission's action plan for a comprehensive Union policy on preventing money laundering and terrorism financing. [s.l.: s.n.], 2020.

EUROPEAN DATA PROTECTION BOARD. Recommendations 01/2021 on the adequacy referential under the Law Enforcement Directive – Adopted on 2 February 2021. 2021.

FERRAZ JÚNIOR, Tércio Sampaio. Comunicação de dados e proteção aos sigilo. In: DONEDA, Danilo; MENDES, Laura Schertel; CUEVA, Ricardo Villas Boas (Coord.). *Lei geral de proteção de dados (Lei nº 13.709/2017)*: a caminho da efetividade. São Paulo: Thomson Reuters Brasil, p. 165-176, 2020.

GRECO, Luís, O inviolável e o intocável no direito processual penal: considerações introdutórias sobre o processo penal alemão (e suas relações com o direito constitucional, o direito de polícia e o direito dos serviços de inteligência). In: WOLTER, Jürgen. *O inviolável e o intocável no direito processual penal*: reflexões sobre dignidade humana, proibições de prova, proteção de dados (e separação informacional de poderes) diante da persecução penal. Madri; Barcelona; Buenos Aires; São Paulo: Marcial Pons, p. 21-82, 2019.

_____; GLEIZER, Orlandino. A infiltração online no processo penal – Notícia sobre a experiência alemã. *Revista Brasileira de Direito Processual Penal*, v. 5, n. 3, p. 1483-1518, 2019.

_____; LEITE, Alaor. Discussão do Supremo sobre caso COAF joga luz em lacuna legislativa. *Folha de S. Paulo*, 19.11.2019. Disponível em: https://www1.folha.uol.com.br/poder/2019/11/discussao-do-supremo-sobre-caso-coaf-joga-luz-em-lacuna-legislativa.shtml. Acesso em: 23 set. 2021.

GLEIZER, Orlandino; MONTENEGRO, Lucas; VIANA, Eduardo. *O direito de proteção de dados no processo penal e na segurança pública*. São Paulo: Marcial Pons, 2021.

GREENLEAF, Graham; TYREE, Alan. Bankers' Duties and Data Privacy Principles: Global Trends and Asia-Pacific Comparisons. In: BOOYSEN, Sandra; NEO, Dora (Org.). *Can Banks Still Keep a Secret?*: Bank Secrecy in Financial Centres Around the World. Cambridge: Cambridge University Press, p. 31-61, 2017.

KALKBRENNER, Arndt; KOCH, Christian. *Bankgeheimnis und Datenschutz*. 4. ed. Wiesbaden: DG Verlag, 2019.

KHAN, Sana. The Fourth AML Directive and the EU's Approach to Data Protection: a precautionary warning. ACAMS Today, 15.07.2016. Disponível em: https://www.acamstoday.org/fourth-aml-directive-eus-approach-to-data-protection/. Acesso em: 23 set. 2021.

LIMA, Renato Sérgio de; BUENO, Samira; MINGARDI, Guaracy. Estado, polícias e segurança pública no Brasil. *Revista Direito GV*, v. 12, n. 1, p. 49-85, 2016.

LUZ, Yuri. Bancos de dados públicos e o compartilhamento com agências penais. In: CRUZ, Francisco Brito; SIMÃO, Bárbara (Ed.). *Direitos fundamentais e processo penal na era digital* [livro eletrônico]. São Paulo: InternetLab, 2021 (sem paginação).

MALISH, Richard. *Financial crime and compliance management under GPR (White Paper)*. [s.l.]: Nice Actimize, 2018.

MENDES, Gilmar Ferreira; BRANCO, Paulo Gustavo Gonet. *Curso de direito constitucional*. 13. ed. São Paulo: Saraiva Educação, 2018 (livro eletrônico).

MENDES, Laura Schertel Ferreira. Autodeterminação informativa: a história de um conceito. *Pensar – Revista de Ciências Jurídicas*, v. 25, n. 04, p. 1-18, 2020.

MORÁN MARTÍNEZ, Rosa Ana. Garantías requeridas en la UE para la transferencia internacional de datos a terceros países en la cooperación judicial penal. In: ARAS, Vladimir Barros; DE MENDONÇA, Andrey Borges; CAPANEMA, Walter Aranha (Org.). *Proteção de dados pessoais e investigação criminal*. Brasília: ANPR, p. 163-196, 2020.

MOTA, Gibran Ayupe; HERKENHOFF, Henrique Geaquinto; LIRA, Pablo *et al*. Constitucionalização da atividade de inteligência – Perspectivas e desafios brasileiros. *Revista Brasileira de Segurança Pública*, v. 12, n. 1, p. 134-150, 2018.

OLIVEIRA, Nina Ribeiro Nery. As novas resoluções do Banco Central e da Comissão de Valores Mobiliários e o sigilo dos dados compartilhados na forma da Lei 9.613/98. *Revista de Direito Penal Econômico e Compliance*, v. 4, p. 162-193, 2020.

PEÑA ZAFRA, Manuel. Vinculación entre protección de datos de carácter personal y prevención de blanqueo de capitales. In: *Estudios sobre controle del fraude fiscal y prevención del blanqueo de capitales*. Navarra: Thomson Reuters Aranzadi, p. 227-239, 2016.

PREZIOSI, Camilleri. Finding the balance between data protection and AML requirements. Lexology. Disponível em: https://www.lexology.com/library/detail.aspx?g=8aabfbf8-33c1-456d-869b-ef1f56ec0e08. Acesso em: 29 set. 2021.

ROGALL, Klaus. Moderne Fahndungsmethoden im Lichte gewandelten Grundrechtsverständnisses. *Goltdammer's Archiv für Strafrecht*, v. 1985, p. 1-27.

_____. Informationseingriff und Gesetzesvorbehalt im Strafprozeßrecht. *Zeitschrift für die gesamte Strafrechtswissenschaft*, v. 103, n. 4, 1991.

_____. § 53. In: WOLTER, Jürgen (Org.). *Kommentar SK-StPO*. [s.l.]: Carl Heymanns, 2018.

SALOMÃO NETO, Eduardo. *Direito bancário*. 3. ed. São Paulo: Trevisan, 2020.

SILVA, Virgílio Afonso da. *Direito constitucional brasileiro*. São Paulo: Editora Universidade de São Paulo, 2021.

SOARES, Hugo. Estupro, dever de comunicação às autoridades e titularidade da ação penal: reflexões derivadas da Resolução do Cremerj nº 296/2019, que estabelece a notificação de estupros aos órgãos competentes investigativos em casos atendidos por médicos no Estado do Rio de Janeiro. In: ESTELLITA, Heloisa;

SIQUEIRA, Flávia (Org.). *Direito Penal da Medicina*. São Paulo: Marcial Pons, p. 347-356, 2020.

SOUTO, Gabriel; ROSAL, Isabela. O direito à proteção de dados pessoais à luz da jurisprudência do STF. *Lapin*, 31.03.2021. Disponível em: https://lapin.org.br/2021/03/31/o-direito-fundamental-a-protecao-de-dados-pessoais-a-luz-da-jurisprudencia-do-supremo-tribunal-federal/. Acesso em: 8 ago. 2021.

SUPREMO TRIBUNAL FEDERAL. RE 377.457, Tribunal Pleno, Rel. Min. Gilmar Mendes, DJe 18.11.2008.

_____. RE 593.727, Tribunal Pleno, Rel. Min. Cezar Peluso, DJ 08.09.2015.

_____. RE 1.055.941, Tribunal Pleno, Rel. Min. Dias Toffoli, J. 04.12.2019, DJe 06.10.2020.

_____. MC-ADIn 6.529, Tribunal Pleno, Relª Min. Cármen Lúcia, DJe 15.10.2020.

_____. ADIn 6.387-MC-Ref, Tribunal Pleno, Relª Min. Rosa Weber, DJe 12.11.2020.

TEIXEIRA, Adriano; WEHRS, Carlos; MADRUGA, Antenor. O valor processual das informações de inteligência financeira obtidas por meio do Grupo Egmont. *JCC*, v. 2, n. 2, p. 21-30, 2019.

VOGEL, Benjamin; MAILLART, Jean-Baptiste (Ed.). *National and international anti-money laundering law*: developing the architecture of criminal justice, regulation and data protection. Cambridge, Antwerp, Chicago: Intersentia, 2020.

WOLTER, Jürgen. *O inviolável e o intocável no direito processual penal*: reflexões sobre a dignidade humana, proibições de prova, proteção de dados (e separação informacional de poderes) diante da persecução penal. São Paulo: Marcial Pons, 2019.

Sobre a autora:

Heloísa Estelita | *E-mail:* heloisa.estellita@fgv.br

Professora da Escola de Direito da Fundação Getúlio Vargas e Coordenadora do Grupo de Ensino e Pesquisa em Direito Penal Econômico e da Empresa na mesma instituição. Bolsista da Alexander von Humboldt Stiftung. Doutora em Direito Penal pela Universidade de São Paulo. Mestre em Direito pela Universidade Estadual Paulista.

Data de submissão: 30 de setembro de 2021.
Data de aceite: 10 de dezembro de 2021.

Panorama sobre o Acesso aos Dados Armazenados em Celular em Situação de Flagrante Delito no Brasil

Overview on Brazil's Access to Data Stored on Cell Phones in a Situation of Flagrant Delicto

AMANDA MATIAS CAVALCANTE DE OLIVEIRA[1]
Universidade Católica de Brasília (UCB).

NÉFI CORDEIRO[2]
Universidade Católica de Brasília (UCB).

RESUMO: O constante desenvolvimento tecnológico dos aparelhos celulares possibilitou a extração de dados que impactam as investigações criminais com relativa facilidade, notadamente nas buscas promovidas pelas autoridades policiais por ocasião de situações de flagrante delito. Partindo da explicação do conceito de dado digital e das distinções dos níveis de tutela dos dados extraíveis dos referidos telefones, o presente trabalho visa traçar um panorama sobre efeitos práticos da distinção entre apreensão e acesso de *smartphones* pelo agente policial em situações de flagrante, impactos nos direitos do suspeito e entendimento jurisprudencial sobre os limites a serem observados em busca da elucidação dos fatos delituosos. Com base nas definições em referências, por meio da revisão da doutrina especializada, dos limites legais impostos pela Constituição Federal, pelo Código de Processo Penal e legislação específica sobre dados digitais, e de precedentes históricos sobre o tema, proferidos tanto pelo Superior Tribunal de Justiça quanto pelo Supremo Tribunal Federal, pretende-se obter um panorama sobre a validade do acesso ao conteúdo armazenado nos aparelhos telefônicos quando houver certeza visual dos crimes, de modo a traçar contornos claros aos agentes policiais das hipóteses e da extensão da visualização do conteúdo dos *smartphones* quando promovidas, sem amparo em decisão judicial prévia, nos casos de prisão em flagrante.

PALAVRAS-CHAVE: Celular; dados; flagrante; apreensão; acesso; garantias.

1 Orcid: https://orcid.org/0000-0001-7142-1767.
2 Orcid: https://orcid.org/0000-0002-1490-3118.

ABSTRACT: The constant technological development of cell phones made it possible to extract data that impact criminal investigations with relative ease, especially in searches carried out by police authorities in cases of flagrante delicto. Departing from the explanation of the concept of digital data and the distinctions in the levels of protection of data extractable from those phones, this paper aims to provide an overview of the practical effects of the distinction between seizure and access of smartphones by police officers in flagrant situations, impacts on suspect's rights and jurisprudential understanding on the limits to be observed in the search for elucidation of the criminal facts. Based on the definitions in references, through the review of specialized doctrine, the legal limits imposed by the Federal Constitution, the Code of Criminal Procedure and specific legislation on digital data, and historical precedents on the subject, issued both by the Superior Court of Justice as for the Federal Supreme Court, it is intended to obtain an overview of the validity of access to content stored on telephone devices in the cases of visual certainty of crimes, in order to draw clear outlines to police officers of the hypotheses and the extent of visualization of the content of the smartphones when promoted, without the support of a previous court decision, in cases of arrest in flagrante delicto.

KEYWORDS: Cell phone; data; flagrant; seizeure; access; guarantees.

INTRODUÇÃO

A versatilidade dos aparelhos de telefonia móvel reconfigurou o estilo de vida da sociedade em uma perspectiva global, reformulando antigos hábitos, criando novas rotinas no âmbito pessoal e profissional e, principalmente, revolucionando a forma de comunicação. Da possibilidade de acesso às notícias, passando por movimentações bancárias e compras a distância, não tardou para que o desenvolvimento tecnológico também fosse adaptado para a prática de toda sorte de delitos.

Diante da popularização dos *smartphones* e ante seu emprego por vezes na execução de delitos cibernéticos, próprios e impróprios, a requisição de acesso aos mencionados dispositivos passou a constituir condição frequente em abordagens policiais: muitas vezes antes mesmo de qualquer indício da prática delitiva.

A apreensão dos aparelhos celulares permite, lícita ou ilicitamente, a possibilidade de acesso investigatório aos contatos, histórico de chamadas, vídeos, fotos, *sites* pesquisados, *e-mails* e mensagens trocadas de forma instantânea por meio de aplicativos, dentre outros tantos elementos úteis à investigação. Surge, então, a necessária discussão dos limites de acesso e de

validade das provas obtidas em dados de celulares: as fronteiras da licitude na apreensão e acesso de celular em atuação policial.

Nessa perspectiva, a compreensão da relevância da prova digital, como representação de um universo de dados nem sempre depreendidos pelos sentidos e somente apreendidos de forma completa por um olhar técnico, permite entender que a possibilidade de acesso ao conteúdo armazenado em um aparelho celular pode representar uma janela para visualizar uma infinidade de elementos: nem sempre úteis à comprovação da hipótese inicial da investigação e, muitas das vezes, exponencialmente invasivos sob a perspectiva do investigado e do seu círculo de convívio.

Repensar a lógica da arrecadação de elementos indiciários de prova sob o viés digital possibilita delimitar o alcance da medida investigativa. Por se tratar de dispositivo informático, que utiliza não apenas as redes tradicionais de telecomunicações, mas, sobremaneira, as redes de internet, tudo o que é acessado no *smartphone* pela autoridade pública, sem autorização judicial, importa em volumoso conjunto de dados.

Dentro do conjunto de informações, a tipologia dos elementos que compõem os dados passíveis de coleta do aparelho celular (base, tráfego e conteúdo), ainda que nem sempre sejam coletados no momento da busca pessoal e da apreensão do dispositivo informático, demonstra a expansão do horizonte de possibilidades de comprovação dos indícios de materialidade e autoria delitiva. Por isso, entender o conceito de dado digital, sua classificação tripartite, a regulamentação no ordenamento jurídico e principalmente o modo como o tema vem sendo abordado pelas Cortes Superiores do país alertam para a necessidade de profunda reflexão sobre o tema no âmbito da academia.

A relevância da pesquisa é demonstrada não apenas pela discussão sobre os elementos que compõem a comunicação, como também para que seja sintetizada a distinção sobre quais garantias constitucionais são aplicadas aos dados armazenados nos *smartphones*. A análise da literatura especializada, ainda escassa, e do repertório de decisões judiciais proferidas pelo Supremo Tribunal Federal e pelo Superior Tribunal de Justiça demonstram que o regramento do acesso às informações disponíveis nos *smartphones* não é claro em asseverar qual é a garantia constitucional a ser acionada nos casos de devassa na situação extrema do flagrante delito ou, ainda, qual meio deve ser empregado para obter seu conteúdo.

Apesar de parecer lógico, a inviolabilidade do sigilo de comunicação telefônica e de dados, prescrita no art. 5º, XII, da Constituição Federal, é limitada ao objeto de sua comunicação e, ainda assim, guarda distinção de tratamento em razão do processo de transmissão ou da finalização do diálogo. E o que se verifica nos precedentes adiante examinados é que a solução encontrada pelos Tribunais Superiores pátrios reside no deslocamento do bem jurídico selecionado para tutela. Agora, a proteção contra a devassa de um aparelho celular não é mais vista em razão do direito à propriedade do bem ou de sua fonte originária de utilização, mas sim pela perspectiva da guarda da privacidade de seu usuário e do consequente respeito à reserva legal.

Com o objetivo de apurar os limites da apuração da verdade no momento das buscas que antecedem à prisão em flagrante, o presente artigo abordará, com apoio na literatura e na jurisprudência dos Tribunais Superiores brasileiros, os limites da atuação do Estado-policial. Por meio da revisão da doutrina especializada sobre o universo da prova digital, a perspectiva constitucional da comunicação travada pelos aparelhos celulares e a ótica do processo penal sobre as diligências cabíveis nas hipóteses de verificação do flagrante descrito, o presente trabalho pretende demonstrar a importância da coleta de dados digitais nos contextos extremos ora apontados.

Para tanto, a primeira seção do artigo será dedicada à compreensão da criação de um marco legal dos dados digitais decorrentes do processo de uso e armazenamento de informações nos aparelhos de telefonia celular, trilhando o histórico do processo, a enunciação do conceito de prova digital, a visão tripartite dos dados nele contidos e os direitos fundamentais relacionados à utilização dos referidos dispositivos.

A segunda seção apresenta o processo de obtenção das provas digitais extraídas de aparelhos celulares no contexto da prisão em flagrante. Nesse ponto, será apresentada uma visão crítica ao processo de abordagem das prisões em flagrante no cenário brasileiro, os limites legais de atuação policial em relação à apreensão e acesso dos dados armazenados nos aparelhos celulares dos suspeitos e um panorama sobre a construção do conceito de prova digital válida a ser obtida no contexto do flagrante.

O tópico final abordará, de forma sucinta, o histórico das principais decisões sobre o tema, o que se entende como passível de acesso sem autorização judicial prévia e os rumos a que se direciona a jurisprudência do Supremo Tribunal Federal na análise de precedente submetido à sistemática

da repercussão geral dos recursos. Ao cabo, pretende-se formular um levantamento sobre os critérios e limites de devassa dos dados armazenados nos *smartphones* pelos agentes policiais, de modo a propiciar a utilização válida e eficiente das informações na investigação e futura instrução processual penal.

1 MARCO LEGAL DA PROVA DIGITAL

1.1 Considerações iniciais sobre o tema

Acompanhado da revolução promovida pela internet e do desenvolvimento do fenômeno da convergência (Barreto, 2020, p. 44), o telefone passou de simples dispositivo de ligações telefônicos entre ausentes para se tornar verdadeiro microcomputador portátil. Nessa transformação, os novos telefones móveis passaram a englobar tanto seu uso original, não mais necessariamente transmitido por redes de telefonia, como também opções de conexão à internet, de acesso a vídeos, fotos, áudios, mapas, dentre outras inúmeras opções constantemente criadas e atualizadas.

Prova disso é o resultado da Pesquisa Nacional por Amostra de Domicílios Contínua, do IBGE – Instituto Brasileiro de Geografia e Estatística (2021, p. 1), que comprovou que, no ano de 2019, o uso do aparelho celular obedecia à seguinte escala de prioridades dentre os entrevistados: 1º) envio e recebimento de mensagens de texto, voz e imagens por aplicativos (95,7%); 2º) conversa por chamada de voz ou vídeo (91,2%); 3º) assistir a vídeos, inclusive de programas, séries e filmes (88,4%); e 4º) envio e recebimento de *e-mails* (61,5%).

A multifuncionalidade do aparelho transformou a vida de todos os seus portadores, que passaram a concentrar, com maior ou menor grau de intensidade, parte considerável de informações pessoais nas memórias dos *smartphones* ou de dispositivos de armazenamento em nuvem instalados no dispositivo e contratados pelo usuário.

Além dos dados nele inseridos, a própria destinação do aparelho, infelizmente, também caminhou para facilitar a execução de toda sorte de delito, aqui englobados aqueles praticados contra a inviolabilidade dos sistemas informáticos (cibercrimes próprios ou puros) e aqueles que podem ser praticados tanto no meio virtual quanto de formas tradicionais (cibercrimes impróprios ou impuros) (Kist, 2019, p. 67-69). Prova disso é a recente publicação da Lei nº 14.155/2021 (Brasil, 2021), que demonstra a preocupação

do legislador ordinário com o agravamento das penas nos casos em que delitos contra o patrimônio ou de fraude forem executados com o auxílio de dispositivos eletrônicos ou informáticos, como os *smartphones* disponíveis no mercado.

O aumento da conectividade dos telefones por meio de serviços de internet disponibilizados por franquias pagas ou por redes sem fio, cada vez mais populares na vida cotidiana, propiciou uma profunda mudança de costumes, a ponto de esses bens se tornarem itens de porte indispensáveis. E, com o aumento da portabilidade, a possibilidade de identificar com maior facilidade os rastros da prática de um crime por meio de um aparelho de fácil manuseio e que contempla tantos dados pessoais tornou o celular um item de extrema valia nas investigações policiais.

Apesar de propiciar o aumento da eficiência dos meios investigatórios, deve haver a preocupação com a criação de limites do acesso estatal à vida privada e à intimidade do titular do aparelho, seja por limites prévios – se possível legais –, seja por controle casuístico da necessidade dessa prova invasiva. Embora portátil e acessível em qualquer revista física, o volume de informações de um celular o diferencia muito de outros objetos de posse pessoal, razão pela qual, na visão de Gloeckner e Eilberg (2019), o acesso ao seu conteúdo demanda não só o recurso à prévia manifestação do Poder Judiciário, e sim à decisão judicial precisa.

Diante da sensibilidade desses dados, o acesso ao aparelho telefônico por terceiro deve ser promovido com cautela tanto da forma de visualização quanto dos limites do que se pode ver e em quais situações. Não pela proteção à propriedade do telefone, mas do que ele contém e até do que representa: a garantia de livre comunicação e de transmissão e guarda de dados, sem interferências de terceiros ou do Estado.

Diante da pluralidade de bens jurídicos a serem tutelados, forçoso reconhecer que a comunicação por meio de telefone, seja ela materializada por voz, imagem ou escrita, possui desníveis no plano constitucional. Numa primeira concepção sobre o tema, o aparelho celular é resguardado pelo sigilo de comunicações telefônica e de dados, aqui compreendida a forma telemática, e passível de quebra por decisão judicial, para fins de prova em investigação criminal ou em instruções penais, na forma da Lei de Interceptações Telefônicas posteriormente editada, por força do prescrito no art. 5º, XII, da Constituição Federal.

Os ditames da Lei nº 9.296/1996, todavia, encontram limite na comunicação em processo de transmissão, que não pode ser captada de outra forma, sob pena de perecimento. E, nesse sentido, importante o registro de que a imposição ao acusado de disponibilização de ligações no modo viva-voz não pode ser confundida com a interceptação autorizada judicialmente. Aliás, esse modo informal de escuta é taxado como ilícito pelo Superior Tribunal de Justiça desde 2017, em prol da proteção contra a autoincriminação, tal como decidido no precedente exarado no Recurso Especial nº 1.630.097/RJ (Brasil, 2017).

Quanto aos demais dados transmitidos ou comunicados pelo aparelho celular, a saída, como proposto por Mendes e Branco (2021), está em repensar o sigilo das comunicações telefônicas como mecanismo a propiciar a liberdade de manifestação do pensamento. Ou, como proposto por Badaró (2010), como "[...] mecanismo de salvaguarda do direito à liberdade de manifestação do pensamento de forma reservada [...]".

E se é possível ponderar que o pensamento externado pelo uso do telefone, na perspectiva do art. 5º, XII, da CF/1988, tutela a comunicação em si, forçoso reconhecer que os demais dados armazenados em aparelhos telefônicos não possuem uma proteção tão clara assim no ordenamento jurídico pátrio. Aliás, em relação às comunicações já finalizadas, armazenadas no aparelho celular e não mais passíveis de interceptação, defendem Antonialli *et al.* (2019) que não há sequer consenso sobre o dispositivo constitucional aplicado, sobre os requisitos de padrão probatório de eventual decisão judicial de quebra de sigilo ou sobre a possibilidade de emprego da busca e apreensão como regime jurídico de obtenção de informações armazenadas nos celulares.

Em razão do eloquente silêncio do legislador constitucional e ordinário, a proteção dos dados disponíveis nos mencionados aparelhos, sejam eles decorrentes de comunicações pretéritas ou de outros elementos relativos ao conteúdo da mensagem, passou a ser amparada pela proteção à privacidade e à intimidade, na forma do art. 5º, X, da Carta Magna, como vem reconhecendo a jurisprudência das Cortes Superiores do País.

E, nesse sentido, defende Costa Júnior (2019, p. 127) que "[...] a cláusula de jurisdição continua necessária diante da inevitável subsunção com o inciso X do mesmo dispositivo [...]". Referida tese é corroborada pela edição do Marco Civil da Internet que, nos termos do art. 7º, III, impôs a exigência de ordem judicial para a quebra de sigilo de comunicações armazenadas,

produzidas ou gravadas em meios eletrônicos intermediados pela internet, forma majoritária de comunicação via *smartphone*.

Sob o ponto de vista constitucional, pois, há nítida distinção entre as garantias da intimidade e da privacidade (inciso X), quando comparadas ao sigilo de comunicações (inciso X), que demanda tanto a reserva jurisdicional (expressão "salvo, no último caso, por ordem judicial") quanto a legal (nos termos do vernáculo "na forma que a lei estabelecer"). Referida diferença implica a concretização do acesso estatal por indispensável lei prévia, não servindo a tanto a direta autorização constitucional (Souza, 2020, p. 409). A lei prévia sempre seria recomendável, mesmo na proteção da intimidade e privacidade, mas sua exigência constitucional se dá tão somente na proteção ao sigilo das comunicações.

Em consonância com esse entendimento, observa Souza que a obrigação de prévia autorização judicial, fruto da reserva constitucional de jurisdição, é limitada pela Constituição Federal de 1988 apenas à proteção da inviolabilidade do domicílio, do sigilo das comunicações e da vedação das prisões arbitrárias, direitos fundamentais elencados como mais relevantes pelo Poder Constituinte Originário. Dessa forma, em prol da reserva constitucional da jurisdição, compete ao Poder Judiciário dar a primeira e a última palavra sobre eventual relativização dos referidos direitos (Souza, 2020, p. 406-407).

No acesso e proteção aos dados digitais armazenados nos aparelhos celulares, passíveis de busca e apreensão, sem os limites típicos de proteção da interceptação telefônica ou telemática, coube à jurisprudência brasileira a definição dos limites e dos requisitos de admissibilidade, inclusive quanto à reserva de jurisdição. Assim surge o exame do dado digital.

1.2 Definição de prova digital

Alerta Doneda (2019, p. 136) que o dado, ainda que possa ser representado por uma informação, não se confunde com este, pois antecede-a, tratando-se de um conhecimento inventado antes mesmo da própria interpretação.

De igual modo, Hoffmann-Riem observa que, no campo da teoria da informação, os dados, compostos de sinais ou símbolos criados e transportados por meios tecnológicos, não possuem um significado em si. Defende-os, todavia, como juridicamente importantes, a ponto de serem tutelados por

leis de proteção específica porque "[...] o significado é atribuído a eles quando entram em um processo de comunicação de informações por um remetente e geração de informações pelo destinatário, ou seja, tornam-se o objeto da comunicação [...]" (Hoffmann-Riem, 2021, p. 13-14).

Em consonância com o posicionamento do doutrinador alemão, é possível compreender a concepção de dado por meio da reprodução do conteúdo do art. 5º, I, da Lei nº 13.709/2018, popularmente conhecida como LGPD – Lei Geral de Proteção de Dados, que descreve o dado pessoal como "[...] informação relacionada a pessoa natural identificada ou identificável [...]" (Brasil, 2021).

Anote-se ainda que a LGPD distingue o dado pessoal, acima descrito, daqueles pessoais sensíveis e anonimizados, nos termos do art. 5º, II e III, da norma em referência. Os primeiros guardam informações de natureza extremamente íntima e que merecem ser resguardados acima de tudo, versando sobre questões de raça, religião, opinião política, orientação filosófica, religiosa e sexual, informações genéticas, de saúde e biometria. Os demais tratam de elementos informativos pessoais protegidos por regras de anonimato, impossibilitando a identificação do sujeito por métodos razoáveis e disponíveis de filtragem.

Demonstrada a relevância do dado originário para o processo de comunicação e delimitado o objeto das informações que compõem os dados pessoais, nos termos da lei brasileira, o conceito de dado digital passa a ser complementado por uma abordagem técnica, que permite entender a especificidade da comunicação pela via digital.

Conforme classificação tripartite apresentada por Kist (2019, p. 110), os elementos de informação digitais são compostos por dados de base, tráfego ou de conteúdo. O primeiro relaciona-se aos elementos cadastrais fornecidos pelo usuário do serviço de telecomunicação e de informações técnicas que permitam sua conexão à rede, como número de acesso, IP, *login* e senha. Diante de sua instrumentalidade, referidas informações possuem menor grau de proteção, sendo passíveis de obtenção por autoridades administrativas ou por órgãos de persecução penal, independentemente de prévia autorização judicial, na forma do art. 10, § 3º, da Lei nº 12.965/2014 e art. 15 da Lei nº 12.850/2013, respectivamente.

Por sua vez, os dados de tráfego são aqueles produzidos de forma automática pelo sistema em razão do processo de transmissão da comunicação, englobando informações sobre os usuários (nome, número, en-

dereço) e sobre a comunicação promovida (duração, horário, volume de dados produzidos, forma de transmissão da mensagem) e localização dos equipamentos utilizados.

Os dados de localização, verdadeira subespécie dos referidos dados de tráfego, ganham especial relevância no campo das buscas e apreensões de provas digitais porque, de acordo com sua precisão, podem certificar a posição geográfica exata do aparelho celular (latitude, longitude, altitude), o sentido de deslocamento e até mesmo a quais estruturas de telecomunicação está se conectando (Santos, 2004, p. 48), a exemplo das ERBs – Estações Radio Base ou dispositivos de conexão *wi-fi* que encontra pelo caminho.

Traçado o panorama técnico dos dados digitais, de forma assertiva, observa Kist que compõem o objeto da comunicação tanto os componentes anteriores tratados, que desempenham elementos funcionais da mensagem transmitida em si, quanto os dados de conteúdo, que materializam não só o diálogo, como também a própria comunicação (Kist, 2019, p. 111-112).

Na escala de gradação da tipologia apresentada, os últimos merecem maior tutela porque, além de identificar emissores e receptores da comunicação, revelam o real teor dessa interação, seja por meio de elementos textuais, imagens, vídeos ou áudios. Diante de sua natureza reveladora, são protegidos, juntamente com os dados de tráfego, pelas regras de inviolabilidade das comunicações e de sigilos profissionais (Santos, 2004, p. 45).

Feitas essas considerações sobre as informações pessoais e técnicas que compõem o que denominamos de dado digital, é possível enunciar a prova produzida nesse âmbito tecnológico como "[...] qualquer tipo de informação que possa ser extraída de sistemas de computadores ou de outros dispositivos digitais e que possa ser usada para ou refutar uma ofensa ou violação de política [...]" (Maras, 2015, p. 76 – tradução livre).

E, dentro dessa proposta, também deve ser considerada válida a enunciação proposta por Thamay e Tamer (2020, p. 33) de prova digital como "[...] meio de demonstrar a ocorrência de um fato ocorrido em meio digital, ou que tem no meio digital um instrumento de demonstração de determinado fato de seu conteúdo [...]".

Mais do que a formalização de um conceito estanque, a enunciação do que seja uma prova digital é importante para a seara processual penal porque permite ao intérprete compreender a complexidade de informações coletadas de dispositivos tecnológicos como os *smartphones*, vistos não

mais como um telefone, como também como microcomputadores que produzem toda sorte de informações, com maior ou menor grau de intimidade.

Assim, por todo o exposto, não se pode limitar referida concepção de prova digital ao arquivo de foto ou à mensagem postada em um aplicativo de conversas, facilmente depreendidas pela simples visualização. Dessa forma, a compreensão da referida prova deve agregar em seu conceito o local de sua extração, importante referência para aferir o grau de autenticidade, integridade e de respeito à cadeia de custódia, assim como a perspectiva plural do dado a ela relacionada, apta a captar tanto o conteúdo quanto informações técnicas de extrema valia para a instrução probatória.

2 A OBTENÇÃO DA PROVA DIGITAL ORIUNDA DE TELEFONE CELULAR NO CONTEXTO DA PRISÃO EM FLAGRANTE

A prisão em flagrante, descrita nos arts. 301 e seguintes do Código de Processo Penal, constitui medida de natureza pré-cautelar, por meio da qual o particular pode e o agente estatal deve, diante da certeza visual da prática de um delito, restringir a liberdade do autor do fato mediante voz de prisão, colocando-o à disposição do Poder Judiciário para posterior análise criteriosa do preenchimento dos requisitos do *periculum libertatis* e do *fumus comissi delicti* (Lopes Júnior, 2021, p. 260).

Essa possibilidade de comprovação direta pelo sentido da visão delimita o conceito de flagrante delito que, nos termos do art. 302 do CPP, depende da certeza visual do crime (Brasil, 1941). E, para instrumentalizar a comprovação dessa fundada suspeita de que um crime foi praticado, faz-se necessária a coleta pela autoridade policial de evidências, independentemente de autorização judicial – com claro intento probatório (reunir elementos) e de segurança na prisão.

Como alerta Lopes Júnior (2021, p. 230), essa concepção abstrata do que seja fundada suspeita, resquício do autoritarismo da época da publicação do Código, acaba por permitir que a autoridade policial aborde e reviste indivíduos sem critérios claros, o que pode permitir direcionamento para alvos específicos, com potencial discriminatório ou de perseguição pessoalizada. Ressalta o autor, inclusive, que referidos abusos, ainda que passíveis de responsabilização na forma da Lei nº 13.869/2019, incredulamente são frequentemente referendados pelo Poder Judiciário.

Sobre a dificuldade de conceituação da fundada razão que autoriza a busca e a apreensão, o Supremo Tribunal Federal já havia debatido a possibilidade de dar-se contorno à atuação policial em casos de flagrante delito por ocasião do julgamento do Recurso Extraordinário nº 603.616/RO (Brasil, 2015). Muito embora o tema tenha sido abordado no contexto de ingresso em domicílio sem autorização judicial em hipóteses de flagrante por crime permanente, no caso submetido à técnica da repercussão geral, a Corte Suprema já havia observado que a atuação do agente estatal movido pelas fundadas razões deve estar adstrita à exigência de justificação da atuação da polícia por meio de controle a ser exercido *a posteriori* pelo Poder Judiciário, sob pena de nulidade da prova e de responsabilidade disciplinar, civil e penal.

Com a popularização da tecnologia, as afrontas verificadas no cotidiano brasileiro ganharam um novo capítulo na discussão da prevalência da eficiência probatória estatal em desfavor do respeito aos direitos fundamentais. Considerando o conhecimento público de que o aparelho de telefonia móvel hoje é carregado por boa parte da população e que seu uso é cada vez mais intenso, não tardou para que o acesso ao seu conteúdo fosse promovido de forma rotineira como pontapé de toda sorte de investigações.

Assim, além da prática da revista nas vestes e nos automóveis dos suspeitos, referendadas pelo Código de Processo Penal, as fundadas suspeitas passaram a ser fundamento para ação do Estado-policial de realizar a devassa de aparelhos telefônicos, analisados no local do possível flagrante sem qualquer critério, direcionando a autoridade policial para uma atuação exitosa apenas após o conhecimento de algum dado digital revelador.

Tal como defendido por Rosa e Oliveira (2020), o elevado grau de possibilidades de levantamento de indícios propiciados pela evolução dos meios tecnológicos, por mais que aproxime a vigilância estatal de provas da verdade real dos fatos no processo penal, acaba por instrumentalizar a ação penal com elementos que seduzem o senso comum pelo conteúdo. Assim, a preocupação com a legalidade da forma de sua obtenção e, consequentemente, dos direitos violados em razão de sua obtenção em situações extremas, inverte a ordem lógica da análise da prova, de modo que a avaliação do conteúdo da prova é promovida sem grandes questionamentos sobre a validade de sua obtenção.

Além do acesso ilícito à privacidade, permite essa busca exploratória o conhecimento de dados de toda espécie, até mesmo de extremada intimida-

de, seja por fotografias e vídeos, seja pelas ligações registradas, seja mesmo por dados outros contidos no aparelho celular. Por isso, diante da sensibilidade das informações tanto do titular do aparelho quanto dos terceiros de seu convívio, tal como leciona Mendes (2019), é mister a adoção de limites claros ao que denomina de buscas subjetivas em dispositivos informáticos: sobretudo para o uso dos referidos elementos e do conhecimento de informações que não guardam relação com o objeto da fundada suspeição policial.

2.1 Da absoluta distinção entre apreensão e acesso aos dados do aparelho celular apreendido em flagrante

Depreende-se da leitura do disposto nos arts. 240 e 244 do Código de Processo Penal que a única forma de comunicação expressamente disciplinada pelo instituto da busca pessoal é a epistolar, passível de arrecadação pela autoridade policial, "[...] abertas ou não, destinadas ao acusado ou em seu poder, quando haja suspeita de que o conhecimento do seu conteúdo possa ser útil à elucidação do fato [...]" (Brasil, 1941), tal como prescreve a alínea *f* do art. 240, § 1º, da mencionada norma.

Muito embora inexista regramento específico na lei processual penal sobre a apreensão de telefones móveis, notadamente daqueles dotados de microcomputadores como os *smartphones* atualmente disponíveis no mercado, não se pode esquecer que o CPP é datado da década de 1940, quando referido serviço de telecomunicação sequer existia. E isso, por si só, demonstra o descaso do legislador com meio probatório tão complexo, relegando para a apreciação do Poder Judiciário a difícil tarefa de lidar com a casuística.

Em consonância com a crítica à deficiência de tratamento da prova digital na legislação processual penal, Gloeckner e Eilberg (2020) ressalvam que a confusão remete à própria inserção do instituto da busca e apreensão no Título VII do Código de Processo Penal, dedicado às provas. A ausência de parâmetros claros sobre os limites de prova e regime jurídico de obtenção de provas digitais, na visão dos autores, "[...] contribui enormemente para a manutenção de desalinhos e desacordos constitucionais e convencionais [...]" (Gloeckner; Eilberg, 2020), notadamente porque o instituto da busca e apreensão, a despeito de estar atrelado ao devido processo legal e ao preenchimento de auto circunstanciado detalhado, hoje se mostra ineficiente para a tutela de todos os direitos fundamentais envolvidos nessa operação.

E isso é facilmente perceptível pelo tratamento do tema na esfera do Poder Judiciário brasileiro, que precisou dirimir conflitos desde os anos 2000 sobre a compatibilidade legal do acesso ao conteúdo armazenado nos aparelhos celulares nas hipóteses de prisão em flagrante com base em interpretações extensivas e analógicas.

A título exemplificativo, a despeito da antiguidade dos dados coletados, relevante pesquisa divulgada no ano de 2019, promovida com base na amostra inicial de 183 decisões proferidas em sede revisional por 10 distintos Tribunais de Justiça do País entre 12.05.2016 e 14.09.2017, concluiu que há um tratamento díspar na jurisprudência sobre as provas extraídas de aparelhos celulares em situações de flagrante. Dentre os 37 julgados em que o acesso ocorreu após a configuração do flagrante, 73% das provas foram consideradas lícitas, 13,5% ilícitas e outras 13,5% não foram analisadas. Já nos casos em que o policial acessou os dados em operação, sem a verificação do estado de flagrância, a jurisprudência restou dividida no percentual de 50% dos 12 casos analisados (Antonialli *et al.*, 2019).

Sobre o tema, não mais se discute que a visualização de informações pessoais, sensíveis ou não, pelo indevido acesso ao conteúdo do aparelho telefônico móvel acarrete danos às garantias da intimidade e da privacidade, dos sigilos de comunicações e de dados e, até numa visão minoritária, ao domicílio digital (Dezem, 2020; Souza, 2020), tal como assegurado no art. 5º, X, XI e XII, da Carta Magna de 1988. Todavia, o ponto principal das nulidades das provas digitais oriundas desse contexto aparenta ser algo que antecede a análise de seu conteúdo: a abordagem fora das hipóteses legais ou, quando amparadas por força da reserva jurisdicional, em limites que a extrapolem.

2.2 Do panorama jurisprudencial sobre o tema

A dificuldade de balizar os limites de obtenção de provas digitais pelo agente policial no momento da apuração do flagrante pode ser visualizada na prática cotidiana e também na própria evolução jurisprudencial acerca do tema. Sem que se pretenda esgotar o histórico das decisões judiciais proferidas pelos Tribunais Superiores pátrios, com apoio nos principais precedentes ecoados pela doutrina especializada e reverberados pelas decisões judiciais do Superior Tribunal de Justiça e do Supremo Tribunal Federal, o presente tópico promoverá uma breve síntese dos caminhos trilhados pela

jurisprudência sobre as provas digitais arrecadadas de aparelhos celulares até a presente data.

Data de 2006 um dos registros mais antigos sobre a dificuldade de distinção entre apreensão e acesso ao conteúdo do aparelho celular. À época, quando a função principal do dispositivo era a comunicação telefônica, ao julgar o *Habeas Corpus* nº 66.368/PA, considerou o Superior Tribunal de Justiça como válido o acesso aos registros das últimas chamadas pela autoridade policial (Brasil, 2006).

A controvérsia sobre o reconhecimento de quebra de sigilo das comunicações telefônicas foi afastada pelo ministro relator, de forma sucinta, sob o argumento de que o conhecimento dos registros telefônicos não se confunde com o do conteúdo das conversas efetuadas pelo aparelho. Pelo contrário, a checagem da lista de chamadas efetuadas e recebidas no aparelho do suspeito derivaria do dever de cautela geral de arrecadação de elementos informativos na forma prescrita no art. 6º, II e III, do Código de Processo Penal (Brasil, 2006). Pelo que se concluiu da análise do referido julgado, a preocupação da Corte restou limitada à definição do que seria um registro telefônico e de seu impacto em eventual reconhecimento da quebra de sigilo telefônico pelo acesso do agente policial às informações sem prévio amparo judicial.

Em contexto fático semelhante, o Supremo Tribunal Federal também concluiu, no ano de 2012, que o acesso aos registros das últimas chamadas de telefones não configuraria quebra de sigilo telefônico, oportunidade na qual a Corte distinguiu a diferença de tutela entre comunicação telefônica e de registros telefônicos. Na ocasião, a Suprema Corte, nos autos do *Habeas Corpus* nº 91.867/PA, observou que o art. 5º, XII, da Constituição Federal visa proteger a comunicação de dados, e não os dados *per si*, sob pena de inviabilidade de qualquer espécie de investigação criminal.

Em seu voto, o Ministro Gilmar Mendes, relator do *writ*, traçou importantes considerações sobre a distinção entre a apreensão do aparelho celular pela autoridade policial, nos termos do art. 6º do CPP, identificado como meio material indireto da prova, e o efetivo acesso às informações contidas no telefone (registros telefônicos) ou àquelas cadastradas na respectiva empresa de telefonia. E simplifica a distinção ao assentar que "[...] o dado, como no caso, mera combinação numérica, *de per si* nada significa, apenas um número de telefone [...]" (Brasil, 2012), razão pela qual não ha-

veria nenhuma violação à intimidade ou à privacidade do agente criminoso que teve o celular vasculhado pela autoridade policial.

Depreende-se da análise dos dois precedentes acima citados, intensamente debatidos pela doutrina especializada, que a construção do emprego da prova digital extraída dos aparelhos celulares era atrelada ao conhecimento pela autoridade pública de registros telefônicos. Isso porque a proteção auferida pela inviolabilidade do sigilo de comunicações, prescrita no art. 5º, XII, da Constituição Federal, era compreendida como forma de tutelar o ato de se comunicar, e não os dados isoladamente considerados. Assim, acessar, buscar e apreender o telefone celular nas hipóteses de prisão em flagrante era visto pela jurisprudência tanto do Superior Tribunal de Justiça como do Supremo Tribunal Federal como providência legitimada pela previsão do art. 6º do CPP.

Com a evolução dos referidos dispositivos, o conhecimento dos números registrados no histórico de chamadas ou na agenda dos celulares, equiparados ao conhecimento de um papel com anotações encontrado nas vestes do suspeito, perdeu espaço para a nova realidade dos aparelhos de telefonia móvel. Aquele celular, antes limitado ao recebimento e efetivação de ligações telefônicas, transformou-se em verdadeiro computador portátil, migrando sua função originária pela popularização do uso de aplicativos de troca de conversas textuais instantâneas, dentre outras funcionalidades hábeis a facilitar a vida cotidiana.

Data de 2016 importante julgado proferido pela Corte Superior de Justiça, que constitui um marco no tratamento do tema da obtenção de provas por meio do acesso ao conteúdo armazenado nos *smartphones*. Conforme consignado pelo relator do Recurso Ordinário em *Habeas Corpus* nº 51.531/RO, a despeito da possibilidade de acesso, os dados contidos no celular dependem necessariamente de prévia autorização judicial motivada, sob pena de violação ao sigilo telefônico e de dados prescritos no art. 5º, X e XII, da Constituição Federal, dos arts. 1º e 5º da Lei de Interceptações Telefônicas, do art. 3º, V, da Lei de Organização dos Serviços de Telecomunicações (Lei nº 9.472/1997) e art. 7º, I, II e III, do Marco Civil da Internet (Brasil, 2016).

Ainda que se trate de decisão datada pelo ininterrupto avanço da tecnologia, considerando especialmente que o uso de mensageiros instantâneos hoje engloba ligações telefônicas via VoIp e também de chamada de vídeo, o paradigma possui extrema relevância por colocar em destaque a

necessidade de proteção dos dados contidos no aparelho celular em contexto semelhante ao atual, sobremaneira por abranger a questão da proteção dos dados como comunicação complexa, que engloba múltiplas funções de uso.

Digno de nota que referido precedente trouxe a lume tormentosa conclusão, anteriormente afastada nos paradigmas já apreciados: a visualização do conteúdo disposto no *smartphone* pela autoridade policial, diante de sua sensibilidade, afronta a garantia constitucional da privacidade e da intimidade e deve ser precedida de autorização judicial para ser considerada válida (art. 5º, X, da CF/1988 e art. 7º do Marco Civil da Internet). Superada, portanto, a limitação da análise do acesso ao aparelho celular àquilo que era registro ou comunicação propriamente dita, com a consequente demarcação da garantia à privacidade como pedra angular da proteção das comunicações e informações armazenadas nos *smartphones*.

Importante registrar que a ordem identificada no referido acórdão era a tentativa de desatrelar o caso concreto à análise promovida pelo Supremo Tribunal Federal nos autos do *Habeas Corpus* nº 91.867/PA (Brasil, 2012) diante da revolução tecnológica que propiciou que o aparelho celular dispusesse de muito mais elementos do que os telefones existentes no ano de 2012. Com o objetivo de situá-lo historicamente, o Ministro Rogério Schietti Cruz ponderou que o elevado repositórios desses aparelhos permitiria a importação da doutrina americana do direito probatório de terceira geração, que trata do regramento de provas tecnológicas altamente invasivas, dependentes de autorização judicial prévia para sua obtenção (Brasil, 2016, p. 16).

Outro ponto de destaque, levantado pela Ministra Maria Thereza de Assis Moura (Brasil, 2016, p. 27), foi a consignação do argumento de que o acesso *incontinenti* ao conteúdo do telefone móvel também visa atender outro parâmetro constitucional: o direito à segurança pública, prescrito no art. 144 da CF/1988. Dessa forma, tal como defendido pela Magistrada, é importante relembrar que a proteção individual do direito à privacidade do suspeito, como todo direito fundamental, não goza de tutela absoluta. Assim, diante do conflito de interesses entre os órgãos de persecução penal e o titular do aparelho, deve ser posto em prática o juízo de ponderação de interesses, ressalvadas situações excepcionais que autorizariam o imediato acesso ao conteúdo do aparelho, tal como ressalvado pela ministra:

> [...] Não descarto, de forma absoluta, que, a depender do caso concreto, caso a demora na obtenção de um mandado judicial pudesse trazer prejuízos concretos à investigação ou especialmente à vítima do delito, mostre-se possível admitir a validade da prova colhida através do acesso imediato aos dados do aparelho celular. Imagine-se, por exemplo, um caso de extorsão mediante sequestro, em que a polícia encontre aparelhos celulares em um cativeiro recém-abandonado: o acesso incontinenti aos dados ali mantidos pode ser decisivo para a libertação do sequestrado. [...] (Brasil, 2016, p. 27)

Ao cabo, a Corte Superior de Justiça concluiu pela ilicitude da prova decorrente da apreensão de aparelho celular pela autoridade policial, promovida por ocasião da prisão em flagrante, sem autorização judicial, por meio da qual diversos dados pessoais foram devassados pelo agente estatal após o acesso do conteúdo do celular e, especialmente, das mensagens travadas por meio do aplicativo WhatsApp.

Por todo o exposto, a partir da publicação do referido precedente do Superior Tribunal de Justiça, verifica-se que a longa jornada da evolução jurisprudencial brasileira caminha para a flexibilização de toda a informação armazenada nos *smartphones*, desde que preenchidos os requisitos de mandado de busca e apreensão ou de autorização de acesso durante a prisão em flagrante pelo usuário. Fora dessas situações, admite-se ainda o acesso imediato quando registrado elemento de urgência, tal como a hipótese de extorsão mediante sequestro em que é preciso localizar, com urgência, a vítima e seu cativeiro, como destacado, *obter dictum*, no julgamento do Recurso Ordinário em *Habeas Corpus* nº 51.531/RO (Brasil, 2016, p. 26).

Digno de nota que o precedente acima apontado permanece atual e passou a direcionar a atuação do Poder Judiciário de todo o País na análise das provas extraídas de dados armazenados em telefones celulares pelas autoridades policiais. Em consonância com esse entendimento, deve ser registrado que, no ano de 2020, a 2ª Turma do Supremo Tribunal Federal proferiu decisão semelhante à acima debatida, concluindo também pela impossibilidade de visualização de conversas armazenadas no aplicativo de mensagens instantâneas WhatsApp que levaram à busca domiciliar e à consequente condenação do paciente nos autos do *Habeas Corpus* nº 168.052/SP (Brasil, 2020).

Tal como promovido pela Corte Superior de Justiça, a necessidade de autorização judicial foi o argumento principal para consignar que o acesso aos dados armazenados no aparelho celular, em específico mensagens

particulares trocadas via mensageiro instantâneo, deve respeitar a garantia fundamental à privacidade e à intimidade. Ressaltou o Ministro Gilmar Mendes à época que, diferentemente do sustentado no precedente de 2012 (*Habeas Corpus* nº 91.867/PA), a publicação do Marco Civil da Internet e as alterações substanciais do contexto fático, sobremaneira a nova visão do telefone como microcomputador portátil, justificam a mutação constitucional na interpretação dos direitos fundamentais prescritos no art. 5º, X e XII, da CF/1988 (Brasil, 2020, p. 12).

Dessa forma, concluiu a 2ª Turma do Supremo Tribunal Federal, com apoio no voto do Ministro Gilmar Mendes, que, a despeito da possibilidade de acesso ao referido conteúdo de forma emergencial pela autoridade policial, em prol da garantia à privacidade e com base no princípio da proporcionalidade, naquele caso concreto, o acesso ao referido material inserido no *smartphone* dependeria de prévia autorização judicial (Brasil, 2020, p. 12).

A despeito da relevante decisão da 2ª Turma do Supremo Tribunal Federal, que expressamente referencia no acórdão a publicação do Marco Civil da Internet como aprimoramento das normas que tutelam dados armazenados nos aparelhos telefones móveis, o advento de normas de tutela de dados digitais nesse período (*v.g.*, Marco Civil da Internet e Lei Geral de Proteção de Dados, em 2014 e 2018, respectivamente) pouco possibilitou avanços no regramento de acesso do conteúdo nas hipóteses excepcionais do flagrante delito.

Aliás, resolvida a questão da observância da reserva jurisdicional, observam Silva e Moura (2020, p. 420) que "[...] não se sedimentou ainda se, na hipótese de necessidade de mandado judicial, seria preciso um de interceptação telefônica ou de busca e apreensão. Este parece mais coerente, ante a capacidade do aparelho celular de apresentar inúmeras funções [...]".

Casuísticas à parte, o desenvolvimento da obrigatoriedade de decisão judicial motivada ou de autorização, nas hipóteses de flagrante, para apreensão e posterior acesso aos dados inseridos nos *smartphones*, evoluiu para a delimitação de quais dados digitais seriam passíveis de serem considerados como provas válidas para a persecução penal. Nesse sentido, de forma ilustrativa, resume a jurisprudência do Superior Tribunal de Justiça (HC 542.293/SP, 2019) que:

> [...] Os dados armazenados nos aparelhos celulares – envio e recebimento de mensagens via SMS, programas ou aplicativos de troca de mensagens,

> fotografias etc. –, por dizerem respeito à intimidade e à vida privada do indivíduo, são invioláveis, nos termos em que previsto no inciso X do art. 5º da Constituição Federal, só podendo, portanto, ser acessados e utilizados mediante prévia autorização judicial, com base em decisão devidamente motivada que evidencie a imprescindibilidade da medida, capaz de justificar a mitigação do direito à intimidade e à privacidade do agente. [...]

O precedente acima destacado, reproduzido de forma explícita em outros tantos julgados da referida Corte Superior de Justiça, descreve, com precisão, o elenco dos dados armazenados no celular mais comumente arrecadados pelo agente policial: troca de mensagens textuais e fotografias. Tal como exposto no item 2.2 do presente trabalho, contudo, um mínimo de conhecimento técnico pela referida autoridade possibilitaria o acesso a muitos outros dados relevantes, como os dados de localização.

A amplitude dos casos concretos permanece, de forma geral, adstrita às informações passíveis de visualização de forma mais rotineira, até pela limitação pertinente ao local e às condições da abordagem policial, que não podem ser tão extensas, inclusive por segurança do próprio agente. Assim, a criação de um precedente objetivo sobre o tema parece se dirigir à elucidação da distinção de efeitos jurídicos de apreensão e do acesso aos dados digitais, balizada pela obrigatoriedade ou dispensa da reserva jurisdicional.

A pacificação do tema ainda pende de resolução, vez que o Agravo em Recurso Extraordinário nº 1.042.075/RJ (Brasil, 2020), submetido à apreciação do Supremo Tribunal Federal pelo rito da repercussão geral, encontra-se suspenso desde 04.11.2020 por força de voto-vista do Ministro Alexandre de Moraes.

O caso que motivou o julgamento versa sobre o acesso ao conteúdo de aparelho celular abandonado no local do delito que possibilitou a identificação e prisão do suspeito após devassa de dados de índole privada, como fotos, agenda de contatos e lista de chamadas efetuadas.

Até o presente momento, duas teses diversas foram fixadas pelo Ministro Relator Dias Toffoli (Tema 977 da Repercussão Geral) e pelo voto-divergente do Ministro Gilmar Mendes, este último acompanhado pelo voto do Ministro Edson Fachin, nos seguintes termos, respectivamente (Brasil, 2020):

> Tema 977: É lícita a prova obtida pela autoridade policial, sem autorização judicial, mediante acesso a registro telefônico ou agenda de contatos de ce-

lular apreendido ato contínuo no local do crime atribuído ao acusado, não configurando esse acesso ofensa ao sigilo das comunicações, à intimidade ou à privacidade do indivíduo (CF, art. 5º, incisos X e XII).

Tese divergente: O acesso a registro telefônico, agenda de contatos e demais dados contidos em aparelhos celulares apreendidos no local do crime atribuído ao acusado depende de prévia decisão judicial que justifique, com base em elementos concretos, a necessidade e a adequação da medida e delimite a sua abrangência à luz dos direitos fundamentais à intimidade, à privacidade e ao sigilo das comunicações e dados dos indivíduos (CF, art. 5º, X e XX).

Pelo cenário acima exposto, a tese inicial é de licitude do acesso ao que foi validamente apreendido, no limite de registros (e não comunicações) ou dados de maior privacidade. Seria no máximo aferível como condição o exame casuístico de proporcionalidade, com a aferição da necessidade e adequação da medida, sempre respeitados dados mais invasivos da intimidade e privacidade.

A tese divergente não distingue a espécie de informações registradas no aparelho celular, todas elas passíveis da proteção de dados e intimidade, a dependerem, assim, da prévia autorização judicial para o acesso. Aqui a tendência seria compreender que, nos casos de prisão em flagrante, a licitude das provas digitais extraídas de aparelhos celulares dependeria do cumprimento de um requisito importante: o acesso ao conteúdo do dispositivo dependerá de prévia ordem judicial fundamentada e específica, proferida com amparo nos elementos do princípio da proporcionalidade (necessidade e adequação da medida), nos direitos fundamentais à intimidade, à privacidade e ao sigilo das comunicações e dados dos indivíduos.

Por fim, tal como alertado por Kist (2019, p. 404-406), a abordagem policial, seja fruto de busca e apreensão ou de busca pessoal decorrente do flagrante, deverá ser formalizada pelo agente público de forma oficial e extensa. Referida condição de validade do ato permitirá tanto o controle da atuação do Estado-investigador quanto da validade da cadeia de custódia e da legalidade do dado digital apreendido pelo Poder Judiciário, seja na fase inquisitorial, no recebimento da denúncia ou queixa ou, por fim, por ocasião da sentença resolutiva da ação penal.

CONSIDERAÇÕES FINAIS

De todo o exposto no presente trabalho, é possível aferir que a preocupação com os limites e as possibilidades legais de acesso ao aparelho ce-

lular, tanto do suspeito em situação de flagrante delito quanto daquele que teve o celular recolhido por força de decisão judicial, carece de contornos mais claros.

Isso porque o telefone móvel, dada a convergência de informações, passou a abarcar dados digitais de distintas espécies, ora submetidos à cláusula constitucional da reserva jurisdicional, ora à reserva legal derivada das normas infraconstitucionais. E, dentro desse contorno, a distinção do nível de tutela dos dados extraíveis do aparelho celular, sejam eles técnicos como os de base ou de tráfego, em fluxo de transmissão ou relativos às comunicações já finalizadas, acaba por criar distorções em situações limites de apuração da responsabilidade penal.

Prova disso é a irrefutável conclusão de que o que se avançou até agora deve ser creditado mais ao esforço interpretativo do Poder Judiciário, sobretudo das Cortes Superiores brasileiras, que buscaram contornar a discussão sobre a limitação da tutela do art. 5º, XII, da CF/1988, sobretudo pela impossibilidade de aplicação da Lei de Interceptações Telefônicas às comunicações finalizadas e armazenadas nos aparelhos celulares.

A solução proposta pelo Superior Tribunal de Justiça no ano de 2016, por ocasião do julgamento do Recurso Ordinário em *Habeas Corpus* nº 51.531/RO, mostrou que os dados digitais padeciam de proteção constitucional para serem efetivamente tutelados pelo Estado brasileiro. Assim, o entendimento de respeito à privacidade como limite à devassa dos dados digitais disponíveis nos referidos dispositivos mostrou-se importante instrumento de defesa dos titulares de telefonia móvel, especialmente pela conclusão de que os dados disponíveis nos telefones celulares não poderiam ser taxados como de livre acesso às autoridades da persecução penal, sem critérios de admissibilidade ou proporcionalidade.

Referida escolha de fundamentação, por sua vez, acompanhou o próprio processo de desenvolvimento tecnológico, abandonando a discussão do condicionamento da validade da prova digital à mera distinção entre registros telefônicos e comunicação e guarda de dados, como consagrado no paradigma do Supremo Tribunal Federal datado de 2012 (*Habeas Corpus* nº 91.867/PA). A própria evolução dos aparelhos tecnológicos e o desuso das ligações telefônicas tradicionais pela sociedade, tal como constatado pela pesquisa do Instituto Brasileiro de Geografia e Estatística (Brasil, 2021, p. 1), encaminharam a discussão para outros patamares constitucionais que,

por óbvio, não podem mais estar adstritos apenas à inviolabilidade do sigilo de comunicação telefônica ou telemática.

Noutro viés, a despeito da publicação de diplomas normativos importantes para a temática da prova digital, como o Marco Civil da Internet (Lei nº 12.965/2014) e a Lei Geral de Proteção de Dados (Lei nº 13.709/2018), a ausência de regramento na legislação processual penal das providências a serem adotadas pelo agente estatal cria um cenário nebuloso para o acesso ao celular durante a medida pré-cautelar da prisão em flagrante.

A possibilidade de apreensão de aparelhos telefônicos móveis e a possibilidade de acesso a seus dados deve ser objeto de diretrizes claras, preferencialmente legais e prévias, sempre seguidas de regramentos administrativos detalhados para conhecimento, treinamento e adequação da ação investigatória nas ações policiais.

O mero risco de dano irreparável à intimidade e privacidade, do usuário do aparelho e de terceiros, bem ressalta a imprescindível cautela na proteção máxima possível, como garantia fundamental absoluta. Perigosa é a pretensão de superação da obrigatoriedade da reserva legal de jurisdição, de modo a permitir a devassa do conteúdo físico do celular no momento do flagrante, mesmo em situações excepcionais e com proporcionalidade – o debate na Suprema Corte será marco definidor do tema.

Ainda que busque atender o direito difuso à segurança pública prescrito no art. 144 da CF/1988, a devassa ilícita e desarrazoada do aparelho celular configura abuso estatal, justificador de responsabilização civil, penal e administrativa. E, pelo conceito de proporcionalidade, inadmissível será, de todo modo, a devassa que esgote todo o histórico particular, sem limite temporal, das comunicações travadas pelo usuário do aparelho.

Assim, como forma de superar o entendimento de hierarquia de comunicações em transmissão ou armazenadas no aparelho, ou mesmo de que a tutela da privacidade e da intimidade, derivadas da reserva legal, sejam inferiores às protegidas pela Lei de Interceptações Telefônicas, vê-se, com bons olhos, a definição ao menos de *standards* probatórios jurisprudenciais que enunciem mais claramente a tutela dos dados armazenados nos *smartphones*.

Conclui-se, portanto, que o universo das provas digitais oriundas de aparelhos celulares ainda possui novos horizontes a serem descobertos, notadamente porque a tipologia dos dados digitais e a própria tecnologia

passam a permitir possibilidade da descoberta de variadas informações de grande valia à eficiência da persecução. É um caminhar de validade probatória do novo, que precisará ser eficiente, mas que não poderá conviver com o desprezo às garantias do íntimo, do privado, do que é dado digital, com proporcionalidade e controle jurisdicional.

REFERÊNCIAS

ANTONIALLI, Dennys; ABREU, Jacqueline; MASSARO, Heloisa; LUCIANO, Maria. Acesso de autoridades policiais a celulares em abordagens: retrato e análise da jurisprudência de tribunais estaduais. *Revista Brasileira de Ciências Criminais*, Brasília, n. 154, p. 177-214, abr. 2019.

BADARÓ, Guilherme. *Interceptação de comunicações telefônicas e telemáticas*: limites ante o avanço da tecnologia. Ano 2010. Disponível em: http://www.badaroadvogados.com.br/gustavo-badaro-interceptacao-de-comunicacoes-telefonicas-e-telematicas-limites-ante-o-avanco-da-tecnologia-ano-2010.html. Acesso em: 21 mar. 2020.

BARRETO, Alessandro Gonçalves. *Cibercrimes e seus reflexos no Direito brasileiro*. Salvador: JusPodivm, 2020.

BRASIL. Constituição da República Federativa do Brasil de 1988. Brasília: Presidência da República, [1988]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/l10406compilada.htm. Acesso em: 15 set. 2021.

_____. Decreto-Lei nº 3.689, de 3 de outubro de 1941. Código de Processo Penal. Brasília: Presidência da República, [1941]. Disponível em: http://www.planalto.gov.br/ccivil_03/decreto-lei/del3689compilado.htm. Acesso em: 18 set. 2021.

_____. Lei nº 9.296, de 24 de julho de 1996. Lei de Interceptações Telefônicas. Brasília: Presidência da República, [1996]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l9296.htm. Acesso em: 16 set. 2021.

_____. Lei nº 12.965, de 23 de abril de 2014. Marco Civil da Internet. Brasília: Presidência da República, [2014]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm. Acesso em: 14 set. 2021.

_____. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados. Brasília: Presidência da República, [2018]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 16 set. 2021.

_____. Lei nº 14.155, de 27 de maio de 2021. Altera o Decreto-Lei nº 2.848, de 7 de dezembro de 1940 (Código Penal), para tornar mais graves os crimes de violação de dispositivo informático, furto e estelionato cometidos de forma eletrônica ou pela internet; e o Decreto-Lei nº 3.689, de 3 de outubro de 1941 (Código de Processo Penal), para definir a competência em modalidades de

estelionato. Brasília: Presidência da República, [2021]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2019-2022/2021/lei/L14155.htm. Acesso em: 17 nov. 2021.

_____. Superior Tribunal de Justiça (5ª Turma). *Habeas Corpus* nº 66.368/PA. Código de Processo Penal. Denúncia formulada com base no acesso ao registro de chamadas do telefone sem autorização judicial. Pacientes: Davi Resende Soares e Lindomar Resende Soares. Impetrado: Câmaras Criminais Reunidas do Tribunal de Justiça do Estado do Pará. Relator: Min. Gilson Dipp, 5 de junho de 2007. Disponível em: https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=200602016074&dt_publicacao=29/06/2007. Acesso em: 26 set. 2021.

_____. Superior Tribunal de Justiça (6ª Turma). Recurso Ordinário em *Habeas Corpus* nº 51.531/RO. Código de Processo Penal. Lei nº 12.965/2014. Lei nº 9.472/1997. Lei nº 9.296/1996. Acesso ao registro de conversas de WhatsApp pela polícia em telefone sem autorização judicial. Recorrente: Leri Souza e Silva. Recorrido: Ministério Público do Estado de Rondônia. Relator: Min. Nefi Cordeiro, 19 de abril de 2016. Disponível em: https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=201402323677&dt_publicacao=09/05/2016. Acesso em: 26 set. 2021.

_____. Superior Tribunal de Justiça (6ª Turma). *Habeas Corpus* nº 542.293/SP. Constituição Federal de 1988. Lei nº 12.965/2014. Lei nº 9.472/1997. Prisão em flagrante. Acesso ao conteúdo de mensagens em telefone sem autorização judicial. Paciente: Jhones de Fátima Oliveira Alves. Impetrado: Ministério Público do Estado de São Paulo. Relator: Min. Rogerio Schietti Cruz, 17 de dezembro de 2019. Disponível em: https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=201903222817&dt_publicacao=19/12/2019. Acesso em: 26 set. 2021.

_____. Superior Tribunal de Justiça (6ª Turma). Recurso Especial nº 1.630.097/RJ. Constituição Federal de 1988. Lei nº 10.792/2003. Código de Processo Penal. Convenção Americana sobre Direitos Humanos. Prova. Conversa travada por função viva-voz do celular. Dúvidas quanto ao consentimento. Autoincriminação. Descoberta inevitável. Recorrente: Ministério Público do Estado do Rio de Janeiro. Recorrido: Marcelo de Azevedo de Freitas. Relator: Min. Joel Ilan Paciornik, 18 de abril de 2017. Disponível em: https://scon.stj.jus.br/SCON/GetInteiroTeorDoAcordao?num_registro=201602602406&dt_publicacao=28/04/2017. Acesso em: 26 set. 2021.

_____. Supremo Tribunal Federal (Tribunal Pleno Virtual). Agravo em Recurso Extraordinário nº 1.042.075/RJ. Licitude de provas decorrentes do acesso por autoridade policial de aparelho celular em situação de flagrante independentemente de prévia autorização judicial. Agravante: Ministério Público do Estado do Rio de Janeiro. Recorrido: Guilherme Carvalho Farias. Relator: Min. Dias Toffoli, 11 de novembro de 2020. Disponível em: http://portal.stf.jus.br/processos/detalhe.asp?incidente=5173898. Acesso em: 26 set. 2021.

_____. Supremo Tribunal Federal (2ª Turma). *Habeas Corpus* nº 168.052/SP. Licitude de provas decorrentes do acesso por autoridade policial de aparelho celular em situação de flagrante independentemente de prévia autorização judicial. Violação de domicílio. Nulidade de provas. Paciente: Rodrigo Ricardo Laurindo. Impetrado: Tribunal de Justiça do Estado de São Paulo. Relator: Min. Gilmar Mendes, 20 de outubro de 2020. Disponível em: https://jurisprudencia.stf.jus.br/pages/search/sjur437471/false. Acesso em: 21 nov. 2021.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*: fundamentos da Lei Geral de Proteção de Dados. 2. ed. rev. e atual. São Paulo: Thomson Reuters Revista dos Tribunais, 2019.

GLOECKNER, Ricardo Jacobsen; EILBERG, Daniela Dora. Busca e apreensão de dados em telefones celulares: novos desafios diante dos avanços tecnológicos. *Revista Brasileira de Ciências Criminais*, v. 156, p. 353-393, jun. 2019.

HOFFMANN-RIEM, Wolfgang. *Teoria geral do direito digital*: transformação digital: desafios para o direito. Rio de Janeiro: Forense, 2021.

KIST, Dario José. *Prova digital no processo penal*. Leme: JH Mizuno, 2019.

LOPES JÚNIOR, Aury Celso Lima. *Direito processual penal*. 18. ed. São Paulo: Saraiva Educação, 2021. *E-book*.

MARAS, Marie-Hellen. *Computer forensics*: cybercriminals, laws and evidence. 2. ed. Burlington: Jones & Bartlett Learning, 2015.

MENDES, Carlos Hélder C. Furtado. Dado informático como fonte de prova penal confiável (?): apontamentos procedimentais sobre a cadeia de custódia digital. *Revista Brasileira de Ciências Criminais*, v. 161, p. 131-161, nov. 2019.

MENDES, Gilmar Ferreira; BRANCO, Paulo Gustavo Gonet. *Curso de direito constitucional*. 16. ed. São Paulo: Saraiva Educação, 2021. *E-book*.

ROSA, Alexandre Morais da; OLIVEIRA, Daniel Kessler. Novas tecnologias probatórias e o papel do julgador no processo penal. *Revista Brasileira de Ciências Criminais*, v. 167, p. 239-261, maio 2020.

SANTOS, Cristina Máximo dos. As novas tecnologias da informação e o sigilo das telecomunicações. *Revista do Ministério Público*, Lisboa: Sindicato dos Magistrados do Ministério Público, n. 25, p. 44-53, jul./set. 2004.

SILVA, Gabriela Buarque Pereira; MOURA, Tâmara. Prisão em flagrante e o acesso aos dados do celular: desafios entre a privacidade e a investigação criminal. In: ARAS, Vladimir Barros *et al.* (Org.). Associação Nacional dos Procuradores da República (Brasil). *Proteção de dados pessoais e investigação criminal*. Brasília: ANPR, p. 399-430.

SOUZA, Rodrigo Telles de. A exigência de autorização judicial para acesso ao conteúdo de telefone móvel apreendido: uma ampliação da garantia à inviolabilidade domiciliar incompatível com o sistema jurídico brasileiro.

In: SALGADO, Daniel Resende; KIRCHER, Luís Felipe Schneider; QUEIROZ, Ronaldo Pinheiro de (Coord.). *Altos estudos sobre a prova no processo penal*. Salvador: JusPodivm, p. 405-429, 2020.

THAMAY, Rennan; TAMER, Mauricio. *Provas no direito digital*: conceito da prova digital, procedimentos e provas digitais em espécie. São Paulo: Thomson Reuters Brasil, 2020.

Sobre a autora e o autor:

Amanda Matias Cavalcante de Oliveira | *E-mail:* amandamatias.jus@gmail.com

Mestre em Direito pela Universidade Católica de Brasília. Especialista em Direito Constitucional pelo Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa. Bacharel em Ciências Jurídicas pelo Centro Universitário de Brasília. Servidora do Ministério Público Federal.

Néfi Cordeiro | *E-mail:* nefi.cordeiro@msn.com

Doutor e Mestre em Direito pela UFPR, com concentração na área criminal. Professor Universitário – Gradução e Pós-Gradução, *Lato* e *Stricto Sensu*. Graduado em Engenharia Civil (PUCPR). Ex-integrante do Ministério Público e Magistratura estadual no Paraná. Juiz e Desembargador

Data de submissão: 30 de setembro de 2021.
Data de aceite: 2 de dezembro de 2021.

Sistemas de Policiamento Preditivo e Afetação de Direitos Humanos à Luz da Criminologia Crítica

Predictive Policing Systems and Affectation of Human Rights in Light of Critical Criminology

ANA JULIA POZZI ARRUDA[1]
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP).

ANA PAULA BOUGLEUX ANDRADE RESENDE[2]
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP).

FERNANDO ANDRADE FERNANDES[3]
Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP).

RESUMO: A utilização de dados pessoais pelo Poder Público tem ocorrido de forma cada vez mais frequente. Nesse sentido, surgem *softwares* dedicados à análise de relatórios policiais, registros de ocorrências e dados indicativos de criminalidade com o objetivo de prever a ocorrência de crimes futuros, seja com relação à localidade, seja com relação a pessoas específicas, atividade denominada policiamento preditivo. A adoção desses mecanismos coloca em discussão questões relacionadas à função do poder econômico na construção da figura do infrator, afetas ao campo da Criminologia Crítica, de modo que possam ser colocadas em xeque garantias fundamentais como a presunção de inocência e a legalidade das investigações (no campo processual), bem como a privacidade e a igualdade (no campo material). Isto posto, tem-se como objetivo analisar as problemáticas relacionadas aos direitos humanos advindas da incorporação dos sistemas de policiamento preditivo pelos órgãos policiais, à luz das contribuições trazidas pela Criminologia Crítica, na medida em que esta corrente evidencia as conexões entre poder econômico e o sistema de justiça criminal. Trata-se de uma pesquisa jurídico-prospectiva, conduzida sob o método dedutivo e de revisão bibliográfica, uma vez que

1 Orcid: http://orcid.org/0000-0001-7021-7983.
2 Orcid: http://orcid.org/0000-0001-8541-7849.
3 Orcid: http://orcid.org/0000-0002-6801-3356.

se parte dos aspectos gerais do sistema jurídico-penal visando compreender como as particularidades de seu funcionamento poderiam ser reforçadas ou mitigadas, sobretudo no que tange à garantia ou violação de direitos fundamentais, pela introdução de novas tecnologias destinadas à análise de dados pelo Estado com a finalidade de atender algum interesse público. Dentre as conclusões, está o reforço do paradigma de dominação e das noções estereotipadas sobre o indivíduo delinquente, de modo a corroborar com a seletividade penal e a manutenção de uma ordem social estratificada.

PALAVRAS-CHAVE: Inteligência artificial; polícia preditiva; direitos humanos; discriminação; Criminologia Crítica.

ABSTRACT: The use of personal data by the public authorities is each time more frequently. Thereby, softwares dedicated to the analysis of police reports, records of occurrences and data indicative of crime emerges, with the objective of predicting future crimes, either in relation to the location or in relation to specific people, an activity called predictive policing. The adoption of these mechanisms brings into discussion issues related to the role of economic power in the construction of the offender's figure, discussion arose by Critical Criminology, having an effect on fundamental guarantees such as the presumption of innocence and legality of investigations (in the field procedural), as well as privacy and equality (in the material field). The objective is to analyze the problems related to human rights arising from the incorporation of predictive policing systems by police agencies, in light of the contributions brought by Critical Criminology, insofar as this current of thought points out the connections between economic power and the criminal justice system. This is a prospective legal research, conducted under the deductive and bibliographical review method, as it starts from the general aspects of the criminal legal system, aiming to understand how the particularities of its functioning could be reinforced or mitigated, especially in terms of concerning the guarantee or violation of fundamental rights, through the introduction of new technologies for the State to analyze data with the purpose of serving any public interest. Among the conclusions are the reinforcement of the domination paradigm and the stereotyped notions about criminals, in order to corroborate with the penal selectivity and maintenance of a stratified social order.

KEYWORDS: Artificial intelligence; predictive policy; human rights; discrimination; critical criminology.

INTRODUÇÃO

O progressivo desenvolvimento tecnológico tem por objetivo superar os erros da racionalidade humana, aperfeiçoando as atividades realizadas em termos de eficiência e eficácia. Nesse sentido, a utilização de dados pessoais permite que os sistemas sejam programados para fins específicos, de modo a, supostamente, conferir maior objetividade às tarefas e previsibilidade na sua execução, o que otimiza o trabalho de muitos profissionais. Todavia, esses valores não podem ser inseridos no funcionamento institucional

à margem do Direito, sobretudo tendo em vista o contexto democrático de proteção às garantias fundamentais.

Especialmente em razão da relevância que a pauta da segurança adquiriu nos últimos anos, nota-se um interesse cada vez maior em garantir a eficiência da atividade policial, inclusive pela incorporação dos novos avanços tecnológicos e modernas técnicas de vigilância. Nesse sentido, são propostos os sistemas de policiamento preditivo, visando prever a ocorrência de crimes e auxiliar no desenvolvimento de estratégias de segurança pública, com base em métodos de vigilância ostensiva.

No campo jurídico-penal, a questão é especialmente problemática, tendo em vista as reiteradas denúncias de seletividade feitas em face das agências executivas do sistema criminal, como são as polícias. Neste contexto, destacam-se as contribuições da Criminologia Crítica, uma vez que, para além da crítica quanto à desigual distribuição do *status* de criminoso na sociedade, foi capaz de revelar a funcionalidade do sistema jurídico-penal para a manutenção das desigualdades e reprodução das hierarquias de poder em determinada realidade histórica.

Com fundamento na argumentação criminológica e em uma leitura historicizada das relações de poder, portanto, questiona-se a forma de introdução e a lógica de funcionamento dos sistemas de policiamento preditivo, tendo em vista as conflitividades que permeiam a lógica de segurança pública na realidade latino-americana. Consequentemente, destaca-se como esses sistemas, operados sob a lógica do conflito, podem contribuir para a afetação de direitos humanos, especialmente no que tange à igualdade, à dignidade da pessoa humana, à presunção de inocência, dentre outros valores fundamentais para o Estado Democrático de Direito.

Nesse sentido, o presente trabalho propõe reflexão sobre a incorporação de ferramentas que, por meio da análise de dados previamente coletados, objetivam atribuir maior eficiência à atuação policial, seja apontando supostos sujeitos delinquentes, seja buscando locais de maior probabilidade de ocorrência de crimes. A problemática reside na adoção às cegas dessas ferramentas, como se imparciais fossem, sem que haja reflexão sobre a reprodução de padrões historicamente estabelecidos e consolidados.

Assim, propõe-se análise do quadro à luz da Criminologia Crítica, por meio do método dedutivo e de revisão bibliográfica, de modo a proceder a um alerta sobre uma situação que coloca em xeque valores caros ao Estado Democrático de Direito, na qual a adoção, de forma irrefletida, de novas

técnicas visando ao incremento da segurança podem ser tentativas frustradas, que, inclusive, potencialmente reforcem as desigualdades sociais.

1 UTILIZAÇÃO DE DADOS PESSOAIS PARA FINS DE INTERESSE PÚBLICO

As informações pessoais da população há muito tempo representam certa utilidade ao Poder Público, seja por razões de eficiência, seja por razões de controle. Nesse sentido, nas últimas décadas, em decorrência de progressos quantitativos e qualitativos no tratamento de dados pessoais[4], tem-se observado um incremento exponencial nessa utilização.

Conforme explica Bioni, a virtualização da informação, que inicialmente era acumulada, armazenada e transmitida em formato de átomos e posteriormente passou a ocorrer em formato de *bits*, marcou, de forma decisiva, a capacidade de acúmulo de informações (Bioni, 2020, p. 6-7). Assim, associadas a um custo cada vez menor, os diversos dispositivos de memória de computador possibilitaram o armazenamento de quantidades de dados antes inimagináveis, de modo a caracterizar a referida mudança quantitativa. Quanto ao progresso qualitativo, diz respeito ao incremento da técnica aplicada ao tratamento de dados pessoais, como a adoção de novos métodos e a utilização de algoritmos, os quais permitem extração de utilidades diversas das informações.

Destacam-se, nesse sentido, duas das primeiras iniciativas da Administração Pública para centralizar e virtualizar as informações que estavam registradas em suas bases de dados, sob a justificativa de atribuir maior eficiência ao serviço administrativo, denominadas caso *National Data Center* e caso Safari (*Système Automatisé pour les Fichiers Administratifs et le Répertoire de Individus*), que ocorreram, respectivamente, nos Estados Unidos da América (EUA), em 1965, e na França, no início da década de 1970. As iniciativas, apesar de representarem um processo aparentemente burocrático, repercutiram negativamente e levantaram discussões acerca dos direitos dos cidadãos sob os dados, bem como relacionadas aos direitos da personalidade e privacidade (Doneda, 2019, p. 159-165).

4 Tratamento de dados pessoais consiste em "toda operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração", nos termos do art. 5º, X, da Lei nº 13.709/2018.

Em sentido análogo, tem-se a emblemática decisão proferida pela Corte Constitucional alemã na década de 1980, no sentido de considerar inconstitucional uma lei de recenseamento, que impunha coleta compulsória de dados pessoais da população para fins administrativos e estatísticos, além de prever o compartilhamento entre os órgãos da Administração Pública e estabelecer multa para aqueles que se recusassem a responder. Assim, a Corte decidiu que o armazenamento ilimitado de dados pessoais da população constitui grave ameaça à personalidade do indivíduo, bem como reconheceu que o direito fundamental ao livre desenvolvimento da personalidade envolve o direito do indivíduo de determinar o fluxo de suas informações na sociedade (Mendes, 2008, p. 47), o que ficou consagrado como autodeterminação informativa.

Os casos mencionados ilustram a tomada de relevância da proteção de dados pessoais em diversos países, em contraponto aos interesses estatais em coletar, armazenar e utilizar essas informações. É certo, no entanto, que a ascensão das discussões expostas não obsta a utilização de informações pelo Poder Público, mas se desdobraram no estabelecimento de critérios para que esse tratamento ocorresse.

Além disso, o desenvolvimento da doutrina dedicada à proteção de dados pessoais não ocorreu de modo homogêneo nos diversos países, assim como o emprego de dados para fins de utilidade pública. Nesse sentido, chama atenção o fato de que as tensões entre tratamento de dados para atender interesse público e proteção dos mesmos dados como forma de tutela de direitos fundamentais, como liberdade, privacidade e intimidade, ainda nos tempos atuais está colocada em questão.

É certo que a coleta e análise de dados da população está diretamente relacionada à compreensão das necessidades dos indivíduos frente ao Estado, bem como aos interesses institucionais de classificação e exercício das prerrogativas estatais, como assistência e previdência social, além do estabelecimento de política migratória e criminal, por exemplo.

Nesse contexto, importa destacar que a ocorrência de ataques terroristas em diferentes partes do mundo situou a pauta de segurança nacional em posição central a ser discutida pelos países (Stalder; Lyon, 2003, p. 77). Os ataques de 11 de setembro de 2001 representam um marco importante, pois ocasionaram o recrudescimento das formas de controle exercidas pelos EUA, baseados em suposta ameaça terrorista. Não é por acaso que, após

o acontecimento, o país passou a compilar dados de pessoas nacionais e estrangeiras de forma massiva[5].

Assim, a existência de uma constante ameaça, bem como a sensação de medo dela decorrente, somada às possíveis soluções derivadas da automatização de dados pessoais, especialmente frente às expectativas de controle e classificação, foram capazes de fomentar a adoção de práticas de vigilância. No aparente conflito entre liberdades civis e segurança nacional, esta passou a sobrepor-se àquelas, por meio de uma vigilância antecipatória, na qual se baseia a "utopia tecnológica de reconhecer e coibir criminosos antes mesmo que eles tenham a chance de cometer seus crimes"[6], emergindo, portanto, categorias de pessoas suspeitas.

Observa-se, assim, que a tecnologia, ora referida como a virtualização dos dados e utilização cada vez mais frequente e menos custosa, deixou de ser uma situação de fato para ser um vetor condicionante da sociedade, assumindo caráter instrumental e utilitarista, como forma de atingir um fim a ela exterior[7]. Assim, na conjuntura em que novos e primorosos mecanismos de controle são criados, "deterioram-se as tradicionais formas de controle social, cujo lugar é assumido, no entanto, por controles mais penetrantes e globais, tornados possíveis pelo tratamento eletrônico das informações" (Rodotà, 2008, p. 95).

No que diz respeito à coleta de dados da população, sob a justificativa de proteção à segurança nacional, importa ressaltar que nem sempre ocorreu de forma transparente e proporcional. Citem-se as revelações feitas por Edward Snowden em 2013, as quais expuseram práticas do governo norte-americano, no sentido de impor que empresas privadas (especialmente de telecomunicações e tecnologia) cooperassem na interceptação de vasta quantidade de dados e metadados[8] da população, sem que houvesse

5 Nesse sentido, "*one shocking aspect of the increasing convergence of databases since 9/11 is the U.S. government's aggressive acquisition of foreign and domestic databases*" (cf. Webb, 2007, p. 86).

6 Tradução livre de "*the techno-utopian goal is to recognize and apprehend criminals before they have a chance to commit their crimes*" (cf. Stalder; Lyon, 2003, p. 90).

7 Stalder e Lyon destacam, nesse sentido, que a utilização de cartões de identificação, por exemplo, para além de serem simples documentos baseados em papéis, representavam sofisticados instrumentos capazes de conectar mais facilmente a bancos de dados remotos e mecanismos de autenticação. Representavam, portanto, uma solução dada pela iniciativa privada, detentora de tecnologia, para as constantes ameaças existentes. A ressalva que se faz, nesse contexto, é que a utilização de cartões de identificação não necessariamente seria uma forma de coibir práticas terroristas, uma vez que apresenta diversas limitações, mas por certo produzem efeitos discriminatórios negativos em face de grupos sociais determinados (cf. Stalder, Lyon, 2003).

8 Metadados são frequentemente definidos como dados sobre dados; podem revelar interlocutores de uma comunicação, localidade, formato, lapso temporal, dentre outros; podem ser definidos, ainda, pelo "quem, quando e onde" de uma comunicação (cf. Dempsey, 2020, p. 193).

qualquer regime de controle para funcionamento em "conformidade com a lei", sendo, portanto, até então desconhecidas[9].

Importa salientar, ainda, que a coleta de dados em larga escala (dentre os quais se destacam dados de tráfego e dados de localização) para fins de fomentar programas de segurança nacional ou por força de lei não é realizada apenas pelo governo norte-americano. Desse modo, James X. Dempsey aponta para a impossibilidade de catalogar todas as práticas de coleta em massa de dados feitas por países ao redor do mundo, ainda que sejam países democráticos, sentido no qual a vigilância em massa coloca em pauta o poder governamental, a responsabilidade corporativa e a privacidade individual (2020, p. 192-193).

A exposição das práticas de coleta massiva de dados pessoais por parte dos Estados não ocasionou seu encerramento, mas forçou o estabelecimento de parâmetros para que ocorresse, em uma tentativa de promover critérios que permitissem vigilância governamental responsável e aquisição de dados eficaz, de modo a prezar pelos direitos humanos e pelo governo democrático. Nesse sentido, são destacados os princípios da legalidade, proporcionalidade e responsabilidade como basilares para a condução de programas de vigilância estatal (Dempsey, 2020, p. 207-210).

Ademais, a coleta massificada para finalidades imprecisas ou não definidas mostra-se problemática; logo, torna-se relevante a delimitação das finalidades para as quais os dados coletados serão utilizados, bem como a limitação do tratamento ao mínimo necessário para a realização das finalidades[10]. Além disso, a simples menção ao interesse público ou à segurança nacional como justificativa para o tratamento de grande volume de dados pessoais pode ser facilmente alvo de abusos por ser extremamente subjetiva.

> [...] há uma distinção importante de direitos humanos entre situações em que um governo exige de uma corporação dados sobre um determinado indivíduo e, por outro lado, situações em que um governo exige divulgação

9 As revelações feitas pelo ex-funcionário da Agência de Segurança Nacional norte-americana tiveram grande repercussão, inclusive em nível internacional, e foram tidas como prática de espionagem em face dos cidadãos norte-americanos e estrangeiros, bem como de autoridades internacionais, como a ex-Presidenta Dilma Rousseff (cf. Brasil, *online*).

10 Finalidade e necessidade constituem alguns dos princípios nos quais o tratamento de dados pessoais deve estar assentado, conforme estabelecido em legislações de proteção de dados pessoais; vide art. 6º, I e III, da Lei nº 13.709/2018 (Lei Geral de Proteção de Dados) e art. 5º, *b* e *c*, do Regulamento (UE) nº 2016/679 (Regulamento Geral de Proteção de Dados).

> em larga escala ou rotineira de dados sobre muitos indivíduos.[11] (Dempsey, 2020, p. 190)

Portanto, considerando as características da sociedade contemporânea, como o risco e a informatização, tornam-se cada vez maiores as tensões existentes entre a utilização de dados pessoais por parte dos governos para fins de combate ao terrorismo e à criminalidade e a proteção de direitos fundamentais dos cidadãos, notadamente a privacidade[12]. Nesse sentido, observa-se, simultaneamente, a tendência de fortalecimento de direitos humanos e a institucionalização de vigilância em massa nos Estados (Dempsey, 2020, p. 200).

Dentre os mecanismos capazes de institucionalizar a vigilância perpetrada pelo Poder Público, encontram-se os sistemas de policiamento preditivo, objeto de análise do presente trabalho, os quais são amplamente adotados no contexto norte-americano. Esses mecanismos consistem na análise de dados previamente coletados, como relatórios policiais e registros de ocorrências, visando prever crimes, traçar estratégias de segurança pública e otimizar a utilização dos recursos disponíveis.

Assim, a análise estatística de dados intenta predizer áreas de maior criminalidade em determinado período (policiamento preditivo baseado no lugar), bem como as pessoas envolvidas como vítima ou ofensor (policiamento preditivo baseado na pessoa), com o objetivo de otimizar o exercício do controle social e mitigar as taxas de criminalidade (Braga, 2020, p. 693). Com relação à análise de dados pessoais na tentativa de predizer futuros infratores, insta salientar a técnica definida como *profiling*, a qual, por meio da análise de dados relacionados a um indivíduo, atua na formação de perfil comportamental.

> [...] os dados pessoais são tratados com o auxílio de métodos estatísticos e de técnicas de inteligência artificial, com o fim de se obter uma "metainformação", que consistiria numa síntese dos hábitos, preferências pessoais e outros registros da vida desta pessoa. O resultado pode ser utilizado para traçar um

11 Tradução livre de "*there is an important human rights distinction between situations where a government demands from a corporation data regarding a particular individual and, on the other hand, situations where a government demands large scale or routine disclosure of data about many individuals*".

12 A doutrina aponta a privacidade como fundamental na consolidação dos direitos da personalidade, uma vez que essencial para a efetivação da autonomia e livre arbítrio; nesse sentido, a proteção de dados pessoais mostra-se como uma continuação por outros meios do direito à privacidade, pois distancia-se do consagrado "direito a estar só" para significar o controle sobre as informações que dizem respeito à pessoa física (cf. Doneda, 2019; Rodotà, 2008).

> quadro das tendências de futuras decisões, comportamentos e destino de uma pessoa ou grupo. (Doneda, 2019, p. 151)

Destaque-se, nesse sentido, a operação de *softwares* amparados em grandes bases de dados que atuam na realização de uma finalidade previamente determinada, na persecução do interesse supostamente legítimo de auxílio às investigações criminais. Todavia, a ressalva que se faz a esses mecanismos refere-se ao risco de produção de resultados tendenciosos e discriminatórios, sendo relevante a análise crítica da situação.

Além disso, importa que toda inovação introduzida no sistema jurídico-penal esteja adequadamente conformada à finalidade do sistema, tendo-se em vista os conhecimentos produzidos por cada uma das ciências que o compõem, quais sejam, política criminal, criminologia e dogmática penal. Assim, no que tange à criminologia, por ter como objeto tradicional o estudo da criminalidade, importa considerá-la na formulação de estratégias direcionadas a essa questão.

2 O *STATUS* DE CRIMINOSO À LUZ DA CRIMINOLOGIA CRÍTICA

Nota-se um atraso e dificuldade de acolhida das modernas teorias sociológicas sobre o crime no âmbito do direito penal, sobretudo pela maioria dos operadores do Direito, tendo em vista a realidade do ensino jurídico no Brasil, em que muitas universidades privilegiam a formação de profissionais técnicos em detrimento de críticos frente às questões sociais. Alessandro Baratta afirma que tal atraso é minimamente compensado em face de outras agências no sistema de controle que estão em interação constante com a realidade criminal e, em sua formação, entram em contato com a sociologia do fenômeno delitivo, como psicólogos ou profissionais da assistência social (Baratta, 2011, p. 157).

A superação desse atraso de forma definitiva parte da ideia de que não é mais possível se conceber um modelo integrado de ciência penal, em que a Criminologia represente uma "ciência auxiliar" ao Direito (Baratta, 2011). É urgente que o sistema jurídico-penal seja concebido dentro de um corte epistemológico que compreenda a interação entre a Criminologia, a Dogmática e a Política Criminal, reconhecendo-as como ramos autônomos, mas interligadas no que diz respeito ao fundamento teleológico que orienta o pensamento sistemático.

Neste sentido, a relevância da referência à Criminologia Crítica se dá em razão de que esta não é uma ciência neutra, mas apresenta-se como uma ciência social "comprometida com a transformação do próprio objeto", porque visa compreender os conflitos a partir da superestrutura, para, então, resolver as contradições visando à satisfação das necessidades individuais e sociais evidenciadas em contextos historicizados (Baratta, 2011, p. 158). É diante dessa proposta que a Criminologia Crítica adota as bases do materialismo histórico-dialético, pois entende ser imprescindível a mediação entre teoria e práxis, sob a ótica dos conflitos aportados às classes subalternas, para efetiva transformação da realidade e emancipação de grupos sociais.

Importa destacar, ainda que muito brevemente, os antecedentes que contribuíram para o desenvolvimento dessa teoria, de forma a evidenciar a construção teórica do conceito de criminalidade, bem como a mudança de paradigma no objeto de estudo da Criminologia. Tais momentos do discurso criminológico foram fundamentais para gradativamente revelar como as relações de poder e o modo de produção capitalista condicionaram as agências de controle do sistema penal, distribuindo, seletivamente, sua atuação perante determinados setores da sociedade.

As origens do surgimento da Criminologia, enquanto disciplina autônoma e científica por excelência, remontam ao século XIX, a partir dos estudos desenvolvidos por, dentre muitos, Lombroso, Ferri e Garófalo. Em que pese serem relevantes para o desenvolvimento da disciplina as discussões sobre o crime e a pena no âmbito da Escola Clássica[13], é apenas na Escola Positiva que ganha formato científico a preocupação com o estudo das causas do crime e da criminalidade. Funda-se aqui o paradigma etiológico do desvio, segundo o qual o crime é um elemento dado da realidade social, para o qual existem explicações relativas às suas causas:

> Na base deste paradigma, a Criminologia (por isto mesmo positivista) é definida como uma ciência causal-explicativa da criminalidade; ou seja, que tendo por objeto a criminalidade concebida como um fenômeno natural, causalmente determinado, assume a tarefa de explicar as suas causas segundo o método científico ou experimento e o auxílio das estatísticas criminais

13 Destaca-se que a Escola Clássica preocupou-se com a limitação e racionalização do poder punitivo, justamente para romper com a espetacularização das penas, representadas pelos suplícios, que tinham por finalidade evidenciar o poder da Igreja Católica e do soberano como representante de Deus na Terra. Nesse sentido, a obra de Beccaria, a qual reuniu as discussões da época em uma série de justificações para a pena com base, representa a defesa de uma política criminal racional fundada nas premissas iluministas e contratualistas (cf. Beccaria, 2012).

> oficiais e de prever os remédios para combatê-la. Ela indaga, fundamentalmente, o que o homem (criminoso) faz e porque o faz. (Andrade, 1995, p. 24-25)

Ainda que análises posteriores viessem a contradizer as teses do positivismo criminológico, sobretudo que relacionavam as causas da criminalidade a fatores biopsicológicos, como tamanho do maxilar, formato craniano, olhos grandes etc. (Lombroso, 2007), por muito tempo o paradigma etiológico imperou na Criminologia como objeto de investigação do fenômeno criminal. E, mesmo que de forma mais dissimulada e com outras roupagens, ainda podem ser encontradas manifestações desse entendimento nos dias atuais, principalmente na atuação de algumas agências formais de controle. São superadas as conclusões lombrosianas referentes ao atavismo, mas permanece o estudo direcionado a investigar as causas da criminalidade.

É apenas com a teoria do *labelling approach* ou etiquetamento social, a qual tem por antecedentes sociológicos a etnometodologia e o interacionismo simbólico[14], que se desloca o centro de análise das causas de criminalidade para as causas de criminalização, principalmente pelas contribuições das referidas abordagens sociológicas, que compreendiam a sociedade em interação com ela mesma e a respectiva produção de sentido na realidade social. Portanto, se é verdade que a produção do conhecimento depende mais de que sejam feitas as perguntas corretas do que as respostas que serão obtidas, a Criminologia não mais se pergunta aqui "quem é criminoso?" ou "como se torna desviante?"; os teóricos do etiquetamento social passaram a perguntar "quem é definido como desviante?", "o que causa esta definição sobre o indivíduo?" ou, ainda, "quando este indivíduo é objeto de tal definição?" (Baratta, 2011, p. 89).

Dessa forma, Shecaira explica que se trata de uma corrente teórica que desloca a questão criminológica "do plano da ação para o plano da reação (dos *bad actors* para os *powerful reactors*), fazendo com que a verdadeira característica comum dos delinquentes seja a resposta das audiências de controle" (2020, p. 257). A partir deste momento, datado dos anos

14 Perspectiva segundo a qual a sociedade não é uma realidade dada no plano objetivo, mas é o produto de uma construção social, obtida por meio de processos de definição e tipificação realizados por indivíduos de diversos grupos. Nesse sentido, "[...] os interacionistas e os etnometodólogos indicam quais são as *regras gerais*, as *regras de base*, a *cultura comum* que determinam, na interação não oficial, a atribuição da qualidade criminal a certas ações e a certos indivíduos, mas não pesquisam as condições que dão a estas regras, a esta *cultura comum*, um conteúdo determinado, e não um outro" (cf. Baratta, 2011, p. 115).

de 1960 nos Estados Unidos, constitui-se o paradigma da reação social em substituição ao paradigma etiológico da criminalidade, para análise da distribuição do *status* de criminoso ou investigação acerca do próprio processo de criminalização.

Portanto, a tese central é a de que o desvio não é ontológico nem intrínseco à conduta realizada, mas uma qualidade atribuída a determinados sujeitos através de processos formais e informais de definição e seleção, os quais condicionam a reação social das agências de controle (Andrade, 2015). A distribuição do *status* de criminoso ocorre de forma desigual entre os estratos da população, assim como a reação do sistema penal se dá em diferentes intensidades a depender do sujeito desviante. Nesse sentido, a conduta criminal não pode ser concebida de forma independente dos processos sociais de definição e rotulação, razão pela qual, a partir de então, não se fala mais em criminalidade, mas sim em criminalização.

A teoria do *labelling approach* foi responsável por denunciar a seletividade do sistema penal e expor o processo de rotulação, o qual se dava com base nas interações sociais, que, por sua vez, eram condicionadas por fatores políticos, econômicos, raciais, culturais, identitários, dentre outros. Assim, a intervenção do sistema penal não teria o declarado efeito reeducativo, mas determinaria a consolidação da identidade desviante por um processo de estigmatização e, consequentemente, levaria que o indivíduo ingressasse em uma verdadeira carreira criminosa (Baratta, 2011, p. 90). Ou, ainda, conforme explica Vera Regina Pereira de Andrade:

> Se criminal é o comportamento criminalizado e se a criminalização não é mais do que um aspecto do conflito que se resolve através da instrumentalização do Direito e, portanto, do Estado por parte de quem é politicamente mais forte, os interesses que estão na base da formação e aplicação do Direito Penal não são interesses comuns a todos os cidadãos, mas interesses dos grupos que têm o poder de influir sobre os processos de criminalização. (2015, p. 213)

A partir desses aportes teóricos, foi possível que se efetivasse a "passagem" para a Criminologia Crítica, a qual, com fundamento no materialismo histórico-dialético, apreende a conflitividade social tendo em vista o enfoque macrossociológico ligado aos processos de acumulação e distribuição do capital no seio da sociedade de classes. Assim, os autores representantes dessa teoria estabelecem uma relação funcional entre os mecanismos

de controle social e a estrutura econômica capitalista em uma perspectiva historicizada, considerando as particularidades de determinados contextos.

Alessandro Baratta identifica duas etapas decisivas para o surgimento da Criminologia Crítica: a primeira, relacionada à mudança quanto ao enfoque do autor do desvio para as condições objetivas, estruturais e funcionais da sociedade; a segunda, relativa à própria superação do paradigma etiológico referente às causas do desvio pelo paradigma da reação social, e a respectiva construção da realidade criminal por meio dos processos de criminalização (2011, p. 160).

É possível afirmar que, com base no paradigma da reação social, a Criminologia Crítica vai além e problematiza o processo de criminalização a partir da forma como são distribuídos o poder e a propriedade na sociedade capitalista, bem como demonstra a relação de funcionalidade que há entre os mecanismos de seleção e a estrutura de desenvolvimento econômico, dentro da respectiva fase capitalista em que se encontre a sociedade. Ou seja, ao introduzir o elemento econômico-político como questão central no estudo da reação social ao crime, evidenciou que o controle social é construído de acordo com as necessidades da ordem capitalista em determinada fase estrutural ou em face de determinada realidade sócio-histórica (Baratta, 2011, p. 164).

Assim, a qualidade criminal é um *status* atribuído aos indivíduos, diferencialmente distribuído perante os diferentes setores da sociedade a partir de um duplo processo de seleção: em um primeiro momento, relaciona-se à seleção dos bens protegidos pela norma penal, bem como das condutas ofensivas a tais bens; posteriormente, seleciona apenas os indivíduos estigmatizados, dentre todos que cometem práticas delitivas, como alvos do sistema jurídico-penal. Ao analisar esse processo de seleção sob a ótica da acumulação do capital, Baratta chega à definição de criminalidade como "um 'bem negativo', distribuído desigualmente conforme a hierarquia dos interesses fixada no sistema socioeconômico e conforme a desigualdade social entre os indivíduos" (2011, p. 164). O autor afirma que:

> A posição precária no mercado de trabalho (desocupação, subocupação, falta de qualificação profissional) e defeitos de socialização familiar e escolar, que são características dos indivíduos pertencentes aos níveis mais baixos, e que na criminologia positivista e em boa parte da criminologia liberal contemporânea são indicados como as causas da criminalidade, revelam ser, antes, conotações sobre a base das quais o *status* de criminoso é atribuído. (Baratta, 2011, p. 165)

Dessa forma, na sociedade capitalista, o desvio é interpretado de duas formas conforme sejam condutas funcionais ou não às relações de produção. À criminalidade das classes subalternas, que decorre de contradições no processo de acumulação de capital, aplica-se a repressão e a reação institucional; por outro lado, quando se trata de condutas ilegais que possam parecer funcionais a este mesmo processo de crescimento econômico (como a criminalidade econômica, a criminalidade ambiental etc.), tem-se altos níveis de tolerância e reações sociais muito reduzidas.

Isto fica ainda mais evidente com as pesquisas que demonstram a cifra oculta da criminalidade de colarinho branco, comprovando que o comportamento criminoso está distribuído em todas as classes sociais, mas a intervenção penal não. E é justamente o aporte da Criminologia Crítica que denuncia os fundamentos das diferentes reações do sistema penal quando se está diante de formas de criminalidade funcionais ao capitalismo financeiro, por oferecer uma visão macrossociológica deste fenômeno:

> A Criminologia Crítica recupera, portanto, a análise das condições objetivas, estruturais e funcionais que originam, na sociedade capitalista, os fenômenos de desvio, interpretando-os separadamente conforme se tratem de condutas das classes subalternas ou condutas das classes dominantes (a chamada criminalidade de colarinho branco, dos detentores do poder econômico e político, a criminalidade organizada, etc.). (Andrade, 2015, p. 217)

É por isso que a Criminologia Crítica deixa de ser uma teoria da criminalidade para se constituir em uma teoria crítica e sociológica do sistema penal como um todo, pois demonstra como as agências de controle social são funcionais para a manutenção da ordem social estratificada. A crítica ao sistema penal, sob a ótica marxista da produção da criminalização em resposta às demandas do capitalismo, leva a três conclusões essenciais para se entender o nexo de funcionalidade do controle, bem como sua via de (des)legitimação.

A primeira diz respeito ao princípio do interesse social, segundo o qual o Direito Penal protegeria igualmente todos os cidadãos e todos os bens essenciais, quando, na realidade, nota-se que a punição de condutas que afetam bens jurídicos se dá em intensidades diferentes conforme o *status* do indivíduo. A segunda, e diretamente relacionada à anterior, contradiz o princípio da igualdade, segundo o qual a lei penal é igual para todos, ao evidenciar que o *status* de criminoso é desigualmente distribuído e inversamente proporcional à condição socioeconômica da respectiva classe social.

A terceira, por fim, denota que a variável da reação criminalizante não diz respeito à danosidade social da conduta ou à gravidade da infração, mas que responde à forma de distribuição do *status* de criminoso (Baratta, 2011, p. 162).

Assim, a crítica à justiça penal burguesa se dá na medida em que esta revela a contradição entre a igualdade formal e a desigualdade substancial dos indivíduos, que se manifesta com relação às chances de serem definidos e controlados como desviantes (Baratta, 2011). Portanto, o direito penal é o direito desigual por excelência, especialmente porque, sob a ideia de igualdade, legitima as posições desiguais dos indivíduos perante o sistema punitivo, as quais reproduzem as relações sociais de produção.

Todavia, a leitura epistemológica da Criminologia Crítica deve ser, sempre, historicizada e, portanto, não pode se furtar ao reconhecimento de como as características da sociedade contemporânea influenciam na questão criminal. Sendo uma teoria datada dos anos 60, desenvolvida, em um primeiro momento, sobretudo no Norte global, especialmente Europa e Estados Unidos, cumpre reproduzir a análise crítica dentro da realidade latino-americana, notadamente no que tange ao Brasil, a partir das respectivas particularidades históricas locais.

A estrutura das relações de poder no contexto da América Latina revela que os criminalizáveis são os "sobreviventes da colonização exterminadora, pelos escombros das civilizações indígenas, dos africanos e seus descendentes, dos cafuzos, mamelucos, polacas, francesas da *belle époque*, gatunos e demais descartáveis" (Batista, 2012, p. 83). É sobre estes, então, que recai a maior intervenção jurídico-penal, reproduzindo os estereótipos e as estigmatizações que orientam algumas práticas de vigilância.

Zaffaroni expõe que a realidade do controle social na América Latina é produto de uma "transculturação protagonizada, primeiro, pela revolução mercantil e, depois, pela revolução industrial" (1991, p. 65), ou seja, foi sempre concebido com base nas necessidades do sistema econômico e orientado para atender aos interesses do poder hegemônico.

Nesse sentido, complementa chamando atenção para o fato de que "nossa região marginal tem uma dinâmica que está condicionada por sua dependência e nosso controle social está a ela ligado" (Zaffaroni, 1991, p. 66). Com isto, Zaffaroni evidencia que o exercício do poder punitivo, no contexto latino-americano, está inserido em uma superestrutura de relações sociais, econômicas e culturais forjadas sobre um paradigma de dominação

e, funcionalmente, reproduz esta hierarquia social, perpetuando as relações de colonização entre centro e periferia, ou norte e sul global, paradigmas que são reforçados por meio do policiamento preditivo baseado no lugar, que será abordado mais à frente.

Nesse sentido, afirma Vera Malaguti Batista (2012) que a questão criminal deve ser compreendida a partir das diferentes demandas por ordem configuradas em uma conjuntura dada pelas necessidades econômicas, sociais e culturais. Na evolução histórica, as demandas por ordem foram estruturadas a partir dos imperativos do poder: no absolutismo, evidenciava-se a necessidade de firmar a autoridade do soberano acima de todos, demonstrando-a por meio do espetáculo do suplício; no período da revolução industrial, a mão de obra excedente precisava ser disciplinada para o regime de trabalho nas fábricas, permitindo a ascensão da burguesia[15]; atualmente, a evolução tecnológica, científica e comunicacional alcançou tamanho desenvolvimento, chegando-se a falar, inclusive, em capitalismo de vigilância e mercados comportamentais[16].

Dessa forma, ao se analisar o funcionamento dos sistemas de policiamento preditivo sob a ótica das contribuições da Criminologia Crítica, cumpre levantar os riscos de tais sistemas, com relação à distribuição do *status* de criminoso perante a sociedade e, em específico, aos indivíduos que são alvos destes sistemas de controle. Em outras palavras, importa questionar quais dados são utilizados para realizar as referidas predições, como são apreciados pelos mecanismos tecnológicos e de que forma isso reflete na sociedade, uma vez que o processo de criminalização é reforçado pelas atuações seletivas das agências do sistema penal.

15 A este respeito, Foucault (2013, p. 270) demonstra como as práticas de vigilância institucional produzem a delinquência que é funcional à manutenção do sistema: "Essa produção da delinquência e seu investimento pelo aparelho penal devem ser tomados pelo que são: não resultados definitivos, mas táticas que se deslocam na medida em que nunca atingem inteiramente seu objetivo".

16 Conforme Shoshana Zuboff, o capitalismo de vigilância consiste na apropriação do superávit comportamental pelas grandes *big techs*, sendo o Google a pioneira (de forma análoga ao que as empresas GM e Ford representaram para o capitalismo no século XX); nesse sentido, os rastros deixados na internet permitem a análise comportamental de cada pessoa, funcionando como subsídio para um mercado que atua sempre no sentido de explorar hábitos e preferências pessoais (denominado mercado comportamental). A autora ainda assevera que "o capitalismo de vigilância reivindica de maneira unilateral a experiência humana como matéria-prima gratuita para a tradução em dados comportamentais", bem como "os verdadeiros clientes do capitalismo de vigilância são as empresas que negociam nos mercados de comportamento futuro" (cf. Zuboff, 2021, p. 22-27).

3 PROBLEMÁTICAS RELACIONADAS AO POLICIAMENTO PREDITIVO

Os sistemas de policiamento preditivo, quando se propõem a prever crimes e a estabelecer estratégias de segurança pública com base em tratamento de dados, representam uma vigilância automatizada – ou, ainda, criminalização automatizada –, porque reproduzem o *modus operandi* dos agentes de controle social, incidindo, primeiramente, sobre as populações em situação de vulnerabilidade. No contexto da sociedade do risco, em que a sensação social de insegurança é fundamento da expansão do direito penal, a frequente ameaça de risco acaba sendo uma oportunidade para que essas práticas se instalem.

Destaca-se, nesse sentido, que a denominada sociedade do risco é frequentemente marcada pelas incertezas que desafiam a racionalidade humana, especialmente ante a incapacidade de prevê-lo e controlá-lo. Cabe ressaltar que os riscos constituem produtos de falhas humanas, procedentes do progresso científico, e primordialmente de onde emerge necessidade por segurança, previsão e controle. Assim, caracteriza-se o paradoxo existente na situação em que os riscos ocasionam o incremento de técnicas que objetivam segurança e controle, mas que concomitantemente acabam por gerar novos riscos (Beck, 2011).

Além disso, especialmente em uma sociedade globalizada, a característica do risco desconhece fronteiras, sendo integrada por ameaças irrestritas e ao mesmo tempo desconhecidas. Nesse sentido, "os principais meios de obter segurança, ao que parece, são as novas técnicas e tecnologias de vigilância, que supostamente nos protegem, não de perigos distintos, mas de riscos nebulosos e informes" (Bauman; Lyon, 2013, p. 95). A indústria da vigilância desenvolve-se, portanto, sob o pretexto de segurança e vigilância serem sinônimos, sentido no qual se afirma que "as inseguranças são um corolário prático das sociedades securitizadas de hoje" (Bauman; Lyon, 2013, p. 101), ao passo que a adoção dessas medidas não necessariamente aumenta o sentimento de segurança, mas por certo potencializa as discrepâncias sociais.

Em se tratando das estratégias de policiamento preditivo propriamente ditas, as quais são justificadas por maior eficiência e melhor gestão de recursos financeiros e de pessoal nas práticas investigativas, cumpre destacar limitações e falhas. Formulados pela iniciativa privada e vendidos com a promessa da neutralizar os preconceitos dos tomadores de decisão humana, poucos são os fornecedores totalmente transparentes sobre a operação

dos sistemas (frequentemente protegidos pelo segredo industrial), sobre os dados efetivamente utilizados ou sobre o regime de responsabilização adotado pelo fornecedor em casos de danos gerados por vieses ou evidências de má conduta (Braga, 2020, p. 693).

O policiamento preditivo baseado no lugar, conforme exposto, propõe-se a prever onde e quando os futuros crimes ocorrerão, de modo a direcionar maior número de policiais para essa localidade[17]. Nesse sentido, assumindo que a distribuição da criminalidade ocorre de forma desigual, o apontamento de "zonas de perigo" seria capaz de deslocar maior policiamento às periferias das cidades e aos locais de residência de grupos vulneráveis. Outra consequência poderia ser o incremento da percepção de crimes de menor potencial ofensivo, decorrente do policiamento ostensivo em locais específicos.

Destaque-se também que os *softwares*, apesar de não necessariamente atuarem da mesma forma e sob a mesma lógica, representam produtos de uma transculturação, apontada anteriormente por Zaffaroni, e podem ser relevantes exemplos de um modelo colonizador de polícia de ocupação, pois representam técnicas de vigilância condicionadas pelo risco e pela sensação de insegurança[18], os quais, por sua vez, são a origem de muitas das demandas contemporâneas. Em que pese a questão do terrorismo não ser tão incisiva na realidade latino-americana como verificado em relação aos países centrais, nota-se que esta transculturação se dá no sentido de que a incorporação dos sistemas de policiamento preditivo poderia reproduzir as relações de desigualdade características desta região. Ou seja, direcionar a atuação policial de forma ainda mais intensiva e ostensiva às regiões de periferia ou às pessoas em situação de maior vulnerabilidade.

O policiamento baseado na pessoa[19], por seu turno, opera na catalogação de dados de registros oficiais (histórico criminal, prisões, situação de liberdades condicionais recebidas, condenações transitadas em julgado, dentre outros) e dados publicamente disponíveis (inclusive dados de mídia

17 Os sistemas PredPol e HunchLab são exemplos de *softwares* pautados em policiamento baseado no lugar (cf. Braga, 2020, p. 696).

18 Nesse sentido, o idealizador do *software* HunchLab, em um texto publicado no *website* da empresa financiadora (Azavea), justifica a venda do sistema para outra empresa privada, afirmando que as ferramentas de policiamento preditivo têm sido utilizadas para endossar uma vigilância abrangente dos cidadãos, o que o autor acredita ser errado (cf. Cheethan, *online*).

19 Exemplos são o *software* Beware da Intrado e listas de calor (*heat list*) utilizadas pela polícia de Chicago (cf. Braga, 2020, p. 699).

social), com a finalidade de estabelecer uma "pontuação de ameaça" aos indivíduos (Braga, 2020, p. 699). É certo que o funcionamento desses sistemas acaba por determinar o modo de atuação da polícia; assim, o tratamento diferenciado da polícia com relação à população ocasiona interferência na autonomia dessas pessoas, bem como sua relação com a sociedade (Braga, 2020, p. 700).

O *profiling* criminal (decorrente do policiamento baseado na pessoa) ainda suscita uma relevante problematização quanto à garantia fundamental de presunção de inocência. Nesse sentido, haveria impeditivo legal de se iniciar procedimento penal no qual exista uma noção preestabelecida acerca de o indivíduo ter cometido a infração, sentido em que se afirma: "Se a presunção de inocência não for estendida ao *profiling* criminal, ela perderá seu papel de princípio norteador na atual era de vigilância ubíqua e de *big data*" (Gless, 2020, p. 4).

Tomam relevância, nesse sentido, as considerações da Criminologia acerca da criminalização primária e secundária. Enquanto a primeira diz respeito à norma penal em abstrato (relacionada aos conceitos e escolhas dos bens jurídicos penalmente relevantes), a segunda refere-se à atuação dos órgãos investigadores e órgãos judicantes. Assim, conforme referido anteriormente, para além de a própria criminalização primária reproduzir os valores das classes sociais de maior poder econômico, por meio da delimitação do conteúdo e não conteúdo da norma penal, a criminalização secundária reflete preconceitos e estereótipos, segundo os quais são atribuídos juízos diversificados sobre condutas análogas a depender da posição social do acusado (Baratta, 2011, p. 176).

Nesse sentido, Baratta aponta para a existência de um código social que regula a aplicação das normas em abstrato por parte das autoridades, havendo dependência causal entre a delinquência secundária e os efeitos gerados sobre a identidade social do indivíduo quando da primeira condenação (2011, p. 179). Portanto, na circunstância em que a coleta massiva de dados pessoais é prática institucionalizada, o desenvolvimento de sistemas de pontuação de ameaça coloca em risco o instituto da "suspeita razoável", na qual deve estar pautada a investigação de um indivíduo, de modo a reforçar e agravar o processo de criminalização secundária seletiva.

Ademais, coloca-se em questionamento a licitude dessas investigações, especialmente ao se considerar que o *profiling* criminal reforça o paradigma de quem deve ser considerado criminoso com base em sua posição

social e não necessariamente com base no delito praticado, de modo a mitigar a garantia fundamental da igualdade. Questiona-se, ainda, a legalidade das provas produzidas, quando pautadas em busca e apreensão injustificadas, por exemplo:

> Se a atividade policial é baseada em algoritmos que dividem a população em grupos tendo em vista determinadas características, o resultado será uma mudança fundamental em nosso sistema jurídico, que passará a ser caracterizado por um aumento de buscas e detenções ilegais, além das consequentes violações à privacidade e à liberdade de todos os cidadãos. (Gless, 2020, p. 5)

Conforme o elucidado pela Criminologia Crítica de que a intervenção punitiva acaba por reproduzir as desigualdades sociais, ao mesmo tempo em que é funcional à acumulação do capital, os sistemas de policiamento preditivo, quando analisados por esta lógica, reforçam a argumentação criminológica. Assim, a vigilância automatizada, tal qual a vigilância policial, ostensivamente recai sobre as classes marginalizadas e sobre os indivíduos estigmatizados, sendo mais uma engrenagem a operar na manutenção da seletividade penal.

O policiamento preditivo, como instrumento de segurança pública, opera dentro do paradigma de conflito, que é a abordagem política dominante no Brasil, segundo a qual o enfrentamento à criminalidade deve ser realizado por meio de ações repressivas, representado, sobretudo, nos próprios termos utilizados pelos meios de comunicação em massa, como "luta contra a criminalidade" ou "guerra às drogas". Essa forma de entender a segurança pública, como garantia da ordem em detrimento da garantia de direitos, é sustentada pela estrutura social, conformada em torno de uma engenharia de punição e vigilância (Andrade, 2013).

Note-se que a introdução de novas tecnologias no âmbito jurídico-penal, portanto, reproduz a lógica segregacionista que reserva a resposta penal aos indivíduos que não interessam à sociedade neoliberal. Ou melhor, interessam na medida em que fornecem dados que podem auxiliar na conformação de práticas de vigilância, e não na promoção da dignidade humana. Neste sentido, há uma "gestão dos indesejáveis" (Casara, 2018) por meio da racionalidade criminal e repressiva, corroborada pelas técnicas de vigilância[20].

20 Virginia Eubanks, nesse sentido, assevera que a adoção de mecanismos responsáveis pela automatização de serviços públicos representa a expansão das punitivistas estratégias de gestão da pobreza; além disso,

Por meio da vigilância, de um lado, intensifica-se a repressão penal, alimentando o sistema de justiça criminal com os estereótipos de criminoso (e vice-versa, ou seja, reforçando o estereótipo pela ação do sistema de justiça criminal); e, de outro, a segregação social e compartimentalização do espaço público, em domínios privados, como condomínios residenciais ou *shopping centers*, assegurando que determinados indivíduos sejam mantidos à distância e impedidos de acessar estes espaços. É, portanto, a exata representação de como o poder punitivo é utilizado para a hierarquização da sociedade, utilizando-se de marcadores sociais como raça, gênero e condição econômica na base de programação dos *softwares* que atuam no sistema de segurança pública.

Diante de todo o exposto, é indiscutível que os sistemas de policiamento preditivo representam uma ameaça ao direito fundamental de igualdade dos indivíduos, uma vez que, contrariamente à promessa de ferramenta neutra e eficaz, podem acabar por reforçar estereótipos sobre a figura do criminoso, em consonância à teoria das carreiras desviantes (Baratta, 2011, p. 179), reforçada pelo processo de criminalização secundária.

Para além disso, considerando que não só os humanos, mas também os sistemas de inteligência artificial, estão sujeitos a incorrer em erro, é necessário refletir sobre a alegação de ilegalidade da persecução penal por parte de indivíduos corretamente indicados como infratores (denominados verdadeiros positivos). Nesse sentido, Gless aponta que, se a intenção é promover a utilização desses mecanismos de forma responsável e visando à eficiência do Sistema de Justiça, há interesse do próprio Poder Público em ir de encontro ao *profiling* discriminatório e à utilização não monitorada desses mecanismos (2020, p. 6).

Nesse contexto, o Tribunal de Justiça da União Europeia firmou entendimento no sentido de que, apesar de as prestadoras de serviços de comunicações terem obrigação legal de conservação dos dados de tráfego e dados de localização, o acesso generalizado pelas autoridades governamentais jamais deveria ser permitido, mas apenas em casos específicos para fins de combate à criminalidade grave e dentro dos limites do estritamente

modelos e algoritmos preditivos que classificam os indivíduos como "de risco" e "problemáticos" prestam-se à condução de vigilância estatal. Essas tecnologias ditas inovadoras seriam responsáveis por esconder a pobreza do público profissional de classe média e dar à nação o distanciamento necessário para fazer escolhas desumanas, as quais afetam diretamente a vida de indivíduos marginalizados (cf. Eubanks, 2018).

necessário, em atenção aos princípios da confidencialidade das comunicações eletrônica e proporcionalidade (TJUE, 2016).

Assim, visando à compatibilização entre preservar direitos fundamentais e atender interesse público, o Comitê Europeu para Proteção de Dados emitiu uma recomendação sobre as garantias essenciais europeias relativas às medidas de vigilância (EDPB, 2020). O documento elenca quatro garantias essenciais, as quais devem ser tratadas de forma complementar, quais sejam: a) o tratamento deve basear-se em regras claras, precisas e acessíveis; b) é necessário demonstrar a necessidade e a proporcionalidade em relação aos objetivos legítimos prosseguidos; c) deve existir um mecanismo de supervisão independente; d) é necessário que os indivíduos disponham de vias de recurso eficazes.

Nota-se que as recomendações colocadas pelo Comitê Europeu para a Proteção de Dados para regular as informações utilizadas pelas autoridades governamentais, restringindo o acesso apenas em face de persecução da criminalidade grave, revelam a possibilidade de que se faça uma ponderação válida entre interesse público e direitos fundamentais. Diante das problemáticas quanto à utilização dos sistemas de policiamento preditivo, especialmente no contexto em que o sistema jurídico-penal é altamente funcional ao modelo econômico capitalista, identifica-se a urgência que se faz presente para que esta política de segurança pública também seja compatibilizada com direitos fundamentais.

Portanto, restam elucidadas as problemáticas relacionadas à coleta e utilização de dados por parte do Poder Público para fins de repressão da criminalidade. Nesse sentido, a melhor contextualização do futuro da coleta de dados massiva dependeria do estabelecimento de critérios claros, objetivos e transparentes, bem como a existência de vias de controle e revisão. Ainda assim, são controvertidas as evidências que apontam efetividade no tratamento de dados em massa para fins de aplicação da lei ou de segurança nacional (Dempsey, 2020, p. 204-205).

CONCLUSÃO

O emprego de novas tecnologias e a utilização de dados diversos para fins de otimização das práticas de investigação criminal e persecução penal constituem realidade posta em diversos países. Conforme exposto, a análise de informações contidas em registros oficiais (como relatórios policiais e boletins de ocorrência), dados abertos (acessíveis pelo público geral,

inclusive em redes sociais), dados de tráfego e dados de localização associados à utilização de *softwares* servem à predição de criminalidade que, em primeira análise, aparentam ser a grande solução da problemática de criminalidade e segurança nacional.

Entretanto, a análise ponderada da situação, especialmente sob as bases da Criminologia Crítica, revela que o policiamento preditivo pode facilmente reproduzir as falhas existentes no sistema jurídico-penal, quais sejam, a seletividade dos indivíduos criminalizados e a contribuição com a manutenção de uma ordem social estratificada, cujos alvos são prioritariamente pessoas negras e em situação social de vulnerabilidade.

Na realidade em que a tecnologia gera incremento nas práticas de vigilância com o objetivo de apaziguar os riscos e os sentimentos de insegurança, observa-se o aumento concomitante da desigualdade substancial dos indivíduos, em prejuízo aos direitos fundamentais. Em se tratando do policiamento preditivo baseado no lugar, observa-se o reforço ao paradigma de dominação, operante sob a lógica do conflito, no qual se vê a atuação da força policial voltada para as periferias das cidades. Quanto ao policiamento preditivo baseado na pessoa, observa-se a adoção de noções estereotipadas sobre quem é o indivíduo delinquente, em detrimento do direito à presunção de inocência.

Todo o exposto coloca em pauta a licitude dessas investigações, bem como das provas produzidas. Outra problemática diz respeito à privacidade dos indivíduos que são alvos das investigações, diante da prerrogativa estatal em tomar esses dados em detrimento de suposto interesse público. Assim, ainda que existam recomendações sobre o uso adequado dessas ferramentas, é necessário advertir sobre a problemática de reforçarem a atuação dos órgãos policiais com base em discriminações.

REFERÊNCIAS

ANDRADE, Vera Regina Pereira de. *A ilusão de segurança jurídica*: do controle da violência à violência do controle penal. 3. ed. Porto Alegre: Livraria do Advogado, 2015.

_____. A mudança do paradigma repressivo em segurança pública: reflexões criminológicas críticas em torno da proposta da 1ª Conferência Nacional Brasileira de Segurança Pública. *Sequência*, Florianópolis, n. 67, p. 335-356, dez. 2013.

_____. Do paradigma etiológico ao paradigma da reação social: mudança e permanência de paradigmas criminológicos na ciência e no senso comum. *Revista Sequência: Estudos Jurídicos e Políticos*, a. 16, n. 30, p. 24-36, jun. 1995.

BARATTA, Alessandro. *Criminologia crítica e crítica do direito penal*: introdução à sociologia do direito penal. Trad. Juarez Cirino dos Santos. 6. ed. Rio de Janeiro: Revan, 2011.

BATISTA, Vera Malaguti. *Introdução crítica à criminologia brasileira*. 2. ed. Rio de Janeiro: Revan, 2012.

BAUMAN, Zygmunt; LYON, David. *Vigilância líquida*. Trad. Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013.

BECCARIA, Cesare. *Dos delitos e das penas*. Trad. Neury Carvalho Lima. São Paulo: Hunter Books, 2012.

BECK, Ulrich. *Sociedade de risco*: rumo a uma outra modernidade. Trad. Sebastião Nascimento. 2. ed. São Paulo: Editora 34, 2011.

BIONI, Bruno Ricardo. *Proteção de dados*: a função e os limites do consentimento. 2. ed. Rio de Janeiro: Forense, 2020.

BRAGA, Carolina. Discriminação nas decisões por algoritmos: polícia preditiva. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e Direito*: ética, regulação e responsabilidade. São Paulo: Thomson Reuters (Revista dos Tribunais), 2. ed., 2020.

BRASIL. Senado Federal. Denúncias de Snowden revelam amplo monitoramento. Disponível em: https://www12.senado.leg.br/emdiscussao/edicoes/espionagem-cibernetica/contexto-a-guerra-nao-declarada/denuncias-de-snowden-revelam-amplo-monitoramento. Acesso em: 23 jun. 2021.

_____. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados (LGPD). Brasília: Presidência da República, 2018. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 30 ago. 2020.

CASARA, Rubens R. R. *Estado pós-democrático*: neo-obscurantismo e gestão dos indesejáveis. Rio de Janeiro: Civilização Brasileira, 2018.

CHEETHAN, Robert Azavea. Why We Sold HunchLab. Publicado em: 23.01.2019. Disponível em: https://www.azavea.com/blog/2019/01/23/why-we-sold-hunchlab/. Acesso em: 4 set. 2021.

DEMPSEY, James X. Privacy and mass surveillance: balancing human rights and government security in the era of big data. In: GONTIJO, Bruno Miranda; LIMA, Henrique Cunha Souza (Org.); PARENTONI, Leonardo (Coord.). *Direito, tecnologia e inovação*. Belo Horizonte/São Paulo: D'Plácido, p. 189-215, 2020.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*: elementos da formação da Lei Geral de Proteção de Dados. 2. ed. São Paulo: Thomson Reuters Brasil, 2019.

EDPB. European Data Protection Board. Recomendações 02/2020 sobre as garantias essenciais europeias relativas às medidas de vigilância. Publicado em 10.11.2020. Disponível em: https://edpb.europa.eu/our-work-tools/our-documents/recommendations/recommendations-022020-european-essential-guarantees_pt. Acesso em: 30 jun. 2021.

EUBANKS, Virginia Eubanks. *Automating inequality*: how high-tech tools profile, police, and punish the poor. New York: St Martin's Press, 2018.

FOUCAULT, Michel. *Vigiar e punir*: história da violência nas prisões. Trad. Raquel Ramalhete. 41. ed. Petrópolis: Vozes, 2013.

GLESS, Sabine. *Policiamento preditivo*: em defesa dos "verdadeiros positivos". Trad. Heloisa Estellita e Miguel Lima Carneiro. *Revista Direito GV*, São Paulo, v. 16, n. 1, jan./abr. 2020.

LOMBROSO, Cesare. *O homem delinquente*. Trad. Sebastião José Roque. São Paulo: Ícone, 2007.

MENDES, Laura Schertel. *Transparência e privacidade*: violação e proteção da informação pessoal na sociedade de consumo. Dissertação (Mestrado em Direito) – Faculdade de Direito da Universidade de Brasília. Brasília, 2008.

RODOTÀ, Stefano. *A vida na sociedade da vigilância* – A privacidade hoje. Organização, seleção e apresentação de Maria Celina Bodin de Moraes. Trad. Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

SHECAIRA, Sérgio Salomão. *Criminologia*. 8. ed. rev., atual. e ampl. São Paulo: Thomson Reuters Brasil, 2020.

STALDER, Felix; LYON, David. Electronic identity cards and social classification. In: LYON, David (Ed.). *Surveillance as social sorting*: privacy, risk, and digital discrimination. Londres: Routledge, 2003.

TJUE. Tribunal de Justiça da União Europeia. Acórdão de 21.12.2016. Processos apensos C-203/15 e C-698/15. Tele2 Sverige e Watson e o. Disponível em: https://eur-lex.europa.eu/legal-content/PT/ALL/?uri=CELEX:62015CJ0203. Acesso em: 30 jun. 2021.

UNIÃO EUROPEIA. Parlamento Europeu e Conselho. Regulamento (UE) nº 2016/679, de 27 de abril de 2016, relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva nº 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Disponível em: http://data.europa.eu/eli/reg/2016/679/oj. Acesso em: 30 abr. 2021.

WEBB, Maureen. *Illusions of security*: global surveillance and democracy in the post-11/9 word. San Francisco: City Lights, 2007.

ZUBOFF, Shoshana. *A era do capitalismo de vigilância*. Trad. George Schlesinger. Rio de Janeiro: Intrínseca, 2021.

ZAFFARONI, Eugenio Raúl. *Em busca das penas perdidas*: a perda de legitimidade do sistema penal. Rio de Janeiro: Revan, 1991.

Sobre as autoras e o autor:

Ana Julia Pozzi Arruda | *E-mail:* ana.arruda@unesp.br

Bolsista Capes. Mestranda em Direito pela UNESP – Universidade Estadual Paulista Júlio de Mesquita Filho – Faculdade de Ciências Humanas e Sociais, *Campus* de Franca (FCHS/UNESP). Advogada.

Ana Paula Bougleux Andrade Resende | *E-mail:* ana.bougleux@unesp.br

Mestranda em Direito pela Universidade Estadual Paulista Júlio de Mesquita Filho – Faculdade de Ciências Humanas e Sociais, *Campus* de Franca (FCHS/UNESP). Pós-Graduada em Direito Digital pela Faculdade CERS. Advogada.

Fernando Andrade Fernandes | *E-mail:* fernando.a.fernandes@unesp.br

Professor Assistente Doutor da Universidade Estadual Paulista Júlio de Mesquita Filho – Faculdade de Ciências Humanas e Sociais, *Campus* de Franca (FCHS/UNESP). Pós-Doutorado em Direito Penal pela Universidade de Salamanca (2011). Doutorado em Direito pela Universidade de Coimbra (2000).

Data de submissão: 27 de setembro de 2021.
Data de aceite: 2 de dezembro de 2021.

Vigilância, Perfilamento e Tratamento de Dados Pessoais no Contexto do Controle Migratório

Surveillance, Profiling and Processing of Personal Data in the Context of Migratory Control

STÉFANI REIMANN PATZ[1]
Universidade Regional Integrada do Alto Uruguai e das Missões (URI).

THAMI COVATTI PIAIA[2]
Universidade Regional Integrada do Alto Uruguai e das Missões (URI).

RESUMO: O artigo tem como objetivos investigar como se dá a utilização de técnicas de perfilamento e tratamento de dados pessoais no ambiente do controle migratório. Para isso, inicialmente se analisam aspectos gerais da importância dos dados pessoais, da sociedade da vigilância, no que consistem as técnicas de perfilamento e nas repercussões das decisões automatizadas. Na sequência, observa-se como tais técnicas são utilizadas no controle migratório britânico e canadense. Depois, analisam-se as iniciativas brasileiras, com destaque ao projeto piloto Embarque Mais Seguro. Por fim, o artigo dedica-se a investigar como se dá o tratamento dos dados pessoais no âmbito do controle migratório brasileiro, observando o disposto na Lei Geral de Proteção de Dados (LGPD) e o Anteprojeto da LGPD Penal. Trata-se de uma análise doutrinária com revisão bibliográfica referenciada que, após exemplificação prática e situando o atual estágio dessas ferramentas nos controles migratórios de alguns Estados, com esteio no método dedutivo, compreende que o uso de tais ferramentas impacta os direitos humanos dos migrantes. Por fim, conclui-se a contribuição com uma perspectiva referente a novos desafios e novas abordagens de regulação ao tratamento dos dados pessoais no contexto migratório.

PALAVRAS-CHAVE: Vigilância; perfilamento; decisões automatizadas; controle migratório.

ABSTRACT: The article aims to investigate the use of profiling and processing techniques for personal data in the migration control environment. For this, we initially analyze general aspects of the

1 Orcid: https://orcid.org/0000-0002-6375-2942.
2 Orcid: https://orcid.org/0000-0001-7123-0186.

importance of personal data, of the surveillance society, which consist of profiling techniques and the repercussions of automated decisions. Next, it is observed how such techniques are used in British and Canadian migratory control. Then, the Brazilian initiatives are analyzed, as highlighted in the Embarque Mais Seguro pilot project. Finally, the article is dedicated to investigating how the processing of personal data takes place in the context of Brazilian immigration control, observing the provisions of the General Data Protection Law (LGPD) and the Penal LGPD Draft. This is a doctrinal analysis with referenced bibliographic review that, after practical examples and situating the current stage of these tools in the migratory controls of some States, based on the deductive method, understands that the use of such tools impacts human rights of migrants. Finally, the contribution is concluded with a perspective regarding new challenges and new regulatory approaches to the processing of personal data in the migratory context.

KEYWORDS: Surveillance; profiling; automated decisions; migration control.

CONSIDERAÇÕES INICIAIS

Pesquisas do Escritório das Nações Unidas sobre Drogas e Crime (UNODC) indicam que, a curto prazo, as restrições de movimentos a partir do alto controle e/ou fechamento de fronteiras terrestres, marítimas e aéreas irão reduzir o movimento migratório. Entretanto, a médio e longo prazo, podem resultar no aumento da migração, que, afetada pela repercussão econômica da pandemia da Covid-19, não possibilitará regularidade aos migrantes e, portanto, poderá aumentar o tráfico de pessoas (UNODC, 2020, s.p.).

Dados de 2020 mostram que existem cerca de 281 milhões de migrantes internacionais. Mais de 1,1 bilhões de pessoas em movimento no mundo. Mais de 82,4 milhões de pessoas forçadas a deixar suas casas devido a conflitos armados, violência generalizada e desastres naturais. Destes, quase 26,4 milhões são refugiados, e quase metade deles tem menos de 18 anos. Neste momento, uma em cada 95 pessoas na Terra foge de suas casas por causa de conflitos e perseguições (ACNUR, 2021; UN, 2021).

Esses dados indicam que uma parcela considerável da população passou e continuará passando por áreas de segurança, questionários, monitoramento por vídeo, escâneres corporais, entre diversas outras formas de vigilância que fazem parte do procedimento de admissão de viajantes e migrantes dentro do território de determinado Estado. Antes, a decisão de permitir ou negar a entrada em um país era tomada apenas por agen-

tes fronteiriços. Hoje, a decisão é apoiada por sistemas automatizados, que perfilam o candidato com apoio de tecnologias de reconhecimento facial e inteligência artificial, por exemplo[3].

O presente texto respalda-se na compressão que viver na contemporaneidade é habitar um mundo em que os algoritmos cada vez mais julgam decisões importantes nas vidas das pessoas. Com o avanço da *big data*[4], houve a ampliação do emprego de sistemas de inteligência artificial e o aprimoramento de sistemas de decisão automatizados[5].

Diante do exposto, o objetivo do artigo é observar como se dá o uso de tecnologias de perfilamento no ambiente do controle migratório e qual é tratamento dos dados pessoais dos migrantes no ordenamento jurídico nacional. Inicialmente, reflete-se sobre o atual contexto da sociedade dos dados, da vigilância e das decisões automatizadas. Na sequência, analisa-se o uso de tecnologias de perfilamento no ambiente do controle migratório do Reino Unido e do Canadá, e, depois, no Brasil[6]. Por fim, investiga-se qual é a disciplina que se aplica ao tratamento dos dados pessoais no contexto do controle migratório brasileiro, com ênfase no disposto na Lei Geral de Proteção de Dados Pessoais (LGPD) e na redação do Anteprojeto da LGPD Penal.

O método de abordagem será o dedutivo e o método de procedimento, o analítico, por meio da pesquisa indireta com a consulta a livros e revistas científicas. Importante notar que não se busca uma análise exaustiva do testemunho histórico, o que, pela densidade da temática, seria inviável. A proposta é tão somente situar o(a) leitor(a) acerca da temática que ocupa a agenda de proteção de dados e a agenda migratória.

---

3 Desde 2018, muitas companhias aéreas começaram a usar a tecnologia de reconhecimento facial a bordo. Só nos EUA, mais de 15 aeroportos criaram sistemas de correspondência facial para ajudar a embarcar os passageiros com mais rapidez e segurança (Thales, 2021).

4 *Big data* é o termo em inglês que descreve o grande volume de dados gerados e armazenados, que podem se estruturados e não estruturados. Dados estruturados são os dados organizados de alguma forma (banco de dados, planilhas eletrônicas, p. ex.) e dados não estruturados são os dados não submetidos a uma organização definida (*website*, mídia, arquivo de texto, p. ex.). Estima-se que apenas 10% dos dados gerados são estruturados (Kaufman, 2019, p. 32).

5 Inúmeros são os exemplos de como decisões antes tomadas por seres humanos agora dependem de modelos de risco preditivo, sistemas automatizados de *ranking* e de elegibilidade. Dentre eles se encontram as possibilidades de concessão de crédito na compra de um automóvel ou mesmo da casa própria, o cálculo do valor dos juros, a seleção de currículos para acesso ao primeiro emprego ou recolocação no mercado de trabalho e até mesmo o direcionamento de uma investigação por fraude ou a escolha de uma determinada região para a ronda policial (Marrafon; Medon, 2019).

6 A escolha dos casos concretos de utilização de ferramentas de inteligência artificial (IA) no contexto migratório pelos governos do Reino Unido e Canadá deu-se: (i) pela relevância dos projetos; (ii) pelo vasto conteúdo bibliográfico disponível nos meios digitais, e (iii) pelo fato de que, a partir dos exemplos, podem-se observar os riscos de tomada de decisões enviesadas com base nos resultados apresentados pelos sistemas.

DADOS, VIGILÂNCIA E DECISÕES AUTOMATIZADAS

A informação sempre foi um elemento crucial para o desenvolvimento humano. Hoje, mais do que nunca. Sedimentada pela evolução tecnológica, a sociedade de informação criou mecanismos capazes de processar e transmitir informações de maneira cada vez mais veloz, ocasionando novas formas de organização social. Se no passado nos organizávamos como uma sociedade "presencial", no momento somos, em grande parte, digitais. Isso implica sociabilidade amplamente medida por tecnologias que fomentam relações pessoais, culturais, mercadológicas, socioeconômicas e até mesmo de vigilância, o que se dá devido ao desenvolvimento exponencial de diferentes tecnologias, com destaque para as chamadas "tecnologias de informação e comunicação" (TIC) (Oliveira, 2021, p. 29).

Nas palavras de Stefano Rodotà, somos "assediados por computadores, espiados por olhos furtivos, filmados por telecâmeras invisíveis. Os cidadãos da sociedade da informação correm o risco de parecer homens de vidro: uma sociedade que a informática e a telemática estão tornando totalmente transparente" (2008, p. 8). Para o autor italiano, estamos assistindo a uma progressiva extensão das formas de controle social, motivadas, sobretudo, por razões de segurança. Trata-se de uma profunda mudança social. A vigilância passa de excepcional a cotidiana, das classes "perigosas" à generalidade das pessoas, do interior dos Estados ao mundo global (2008, p. 9).

Na visão de Samuel Oliveira, algumas questões preocupantes decorrem do emprego maciço das inovações tecnológicas, envolvendo, por exemplo, "a consolidação da vigilância, a erosão da confiabilidade nos governos e nas instituições e as violações sistemáticas a direitos fundamentais" (2021, p. 29-30). Não há dúvidas de que a vigilância existe há muito tempo. Modos mais antigos, menos formais e menos técnicos de vigilância já existiam quando as pessoas observavam umas às outras dentro do ambiente familiar, religioso, estudantil e até em pequenas cidades.

A vigilância, entretanto, não era institucionalizada – pelo menos não nos momentos iniciais. No entendimento de Samuel Oliveira, novas formas de vigilância surgiram à medida que as instituições mudaram e se tornaram menos centrais diante de uma população crescente, urbanizada, globalizada e móvel. Com o início da era da informação e com a disseminação de instrumentos tecnológicos – computadores, aparelhos celulares, dispositivos vestíveis, entre outros –, a preços acessíveis, a vigilância adquiriu uma dimensão altamente tecnológica (2021, p. 31).

Mas por que o termo "sociedade da vigilância"? Para Samuel Oliveira, o termo deve ser utilizado porque, virtualmente, todas as atividades sociais, institucionais e negociais que possuem alguma relevância na sociedade envolvem a coleta e o monitoramento sistemático de dados, bem como a análise desses dados com o objetivo de tomar decisões, minimizar riscos, classificar grupos sociais e exercer poder (2021, p. 81). A vigilância é, portanto, "monitorar as pessoas a fim de regular ou governar seu comportamento" (Gilliom; Monahan, 2013, p. 9).

Uma das maneiras pelas quais a vigilância tem se concretizado é por meio da utilização massiva de tecnologias de reconhecimento facial (TRF) para fins de segurança e controle. Os exemplos são inúmeros. Na China, 200 milhões de câmeras compõem um sistema de vigilância capaz de identificar basicamente qualquer um dos 1.4 bilhões de habitantes do país (Mozur, 2018, s.p.). Já em Dubai, capital dos Emirados Árabes Unidos, um gigante "túnel-aquário", localizado no principal aeroporto da cidade, conta com mais de 80 câmeras que escaneiam o rosto das pessoas à medida que caminham por ele; realizada a análise das imagens obtidas, o sistema de segurança ou permite que a pessoa ingresse livremente no país ou emite um alerta, indicando a necessidade de uma análise mais aprofundada acerca de sua liberação (ONG, 2017, s.p.).

Para Stefano Rodotà, os riscos da sociedade da vigilância ligam-se tradicionalmente ao "uso político de informações para controlar os cidadãos [...], a ideia de vigilância invade cada momento da vida e se apresenta como um traço próprio das relações de mercado, cuja fluidez diz respeito à possibilidade de dispor livremente de um conjunto crescente de informações" (2008, p. 113). O uso político das informações de que fala o autor italiano pode ser facilmente relacionado à temática do texto, considerando que diversos Estados utilizam informações para aceitar ou negar a entrada de migrantes por meio do controle de fronteiras que perfilam os candidatos, com auxílio de tecnologias de reconhecimento facial e inteligência artificial.

Neste contexto, Rodotà afirma que se materializa a imagem do "homem de vidro", o verdadeiro cidadão desse novo mundo. Uma imagem que, não por acaso, "provém diretamente do tempo do nazismo e que propõe uma forma de organização social profundamente alterada, uma espécie de transformação irrefreável da 'sociedade da informação' em 'sociedade da vigilância'" (2008, p. 113). Para Daniel Solove, o processo de agregação de dados sobre alguém cria uma "pessoa digital: um retrato composto de fragmentos de informações combinadas" (2008, p. 125). Trata-se, portanto,

de um verdadeiro quebra-cabeça digital sobre uma determinada pessoa, um perfil com base nos dados disponíveis nos sistemas sobre a pessoa.

No entendimento de Rodotà, a utilização de informações pessoais para construção de perfis individuais ou de grupos deve ser observada com cuidado, considerando que:

> [...] as informações utilizadas são, de fato, sempre parciais e incompletas, mesmo quando se recorre a uma multiplicidade de bancos de dados. Além disso, permanece controversa, e a ser comprovada, a plena validade científica dos modelos usados para produzir novas informações (perfis ou outras) com base em dados coletados. Chega-se assim a "metaconhecimento" sobre as pessoas, que dificilmente podem ser verificados pelos interessados, embora até embasem decisões sobre eles. (2008, p. 115)

A técnica do perfilamento (*profiling*) está presente nos mais diversos contextos e pode ser utilizada com as mais diversas finalidades. A criação de perfis é um processo pelo qual se busca descobrir correlações entre dados que podem ser usados para identificar e representar um indivíduo ou grupo. Esses dados são transformados em conhecimento ou inferências, que, por sua vez, são utilizados para individuar e representar um sujeito ou para identificar um sujeito como membro de um grupo ou categoria, com a construção de prováveis atributos ou comportamentos de uma pessoa (Hildebrandt, 2008).

Nesse sentido, o Regulamento Geral sobre Proteção de Dados (RGPD)[7] da União Europeia define *profiling* em seu art. 4º, item 4, a saber:

> [...] qualquer forma de tratamento automatizado de dados pessoais que consista em utilizar esses dados pessoais para avaliar certos aspetos pessoais de uma pessoa singular, nomeadamente para analisar ou prever aspetos relacionados com o seu desempenho profissional, a sua situação económica, saúde, preferências pessoais, interesses, fiabilidade, comportamento, localização ou deslocações. (União Europeia, 2016)

Ao interpretar o art. 4º do RGPD e registrar que o processo de definição de perfis pode implicar um conjunto de deduções estatísticas, o extinto

7 O Regulamento Geral sobre Proteção de Dados (RGPD) – *General Data Protection Regulation* (GDPR) passou a ser aplicável a partir de 25 de maio de 2018, em substituição à legislação europeia acerca da proteção de dados (substituiu a Diretiva de Proteção de Dados de 1995 – 95/46/EC). A GDPR foi projetada para: a) harmonizar as leis de privacidade de dados em toda a Europa; b) proteger e capacitar a privacidade de dados de todos os cidadãos da UE; e c) remodelar a maneira como as organizações em toda a região abordam a privacidade dos dados.

Grupo de Trabalho do Artigo 29[8] esclarece que o *profiling* é frequentemente "utilizado para efetuar previsões sobre as pessoas, recorrendo a dados provenientes de várias fontes para inferir algo sobre uma pessoa, com base nas qualidades de outras pessoas que, estatisticamente, parecem semelhantes" (European Commission, 2017, p. 7).

Para Felix Naumann, o processo de geração de perfil de dados, ou *data profiling*, consiste em examinar os dados disponíveis em uma determinada fonte e coletar informações a respeito deles, produzindo metadados cuja análise é um passo importante para gerenciar a qualidade dos dados da fonte. Um cenário típico seria a varredura das tabelas de um banco de dados relacional para obter informações como tipos de dados, padrões de valores, completude e unicidade de colunas, ou até mesmo dependências funcionais e regras de associação (2014, p. 40-49).

Cabe pontuar, entretanto, que a geração de dados, além das bases de dados relacionais tradicionais, cresce na atualidade tendo em vista o volume de dados gerados a cada instante, algo sem precedentes na História. Nesse sentido, dados do Banco Mundial indicam que 3,000,000,000,000 *gigabytes* (GB) circulam globalmente através da internet. Isso equivale a 32 GB por mês e, por pessoa de dados, mais de 1 GB por dia, ou 325 milhões de casas assistindo a vídeos em *streaming* simultaneamente (2021, s.p.).

A título exemplificativo, Danilo Doneda indica que o perfilamento poderia ser utilizado para, por exemplo, controlar a entrada de pessoas em um determinado país pela alfândega e poderia ser utilizada pelas empresas para traçarem perfis de consumidores e, assim, direcionarem a publicidade (2006, p. 173).

Neste cenário, é preciso lembrar que a incorporação de sistemas de decisões automatizados que empregam inteligência artificial (IA) tem sido adotada com entusiasmo pelo setor privado e público em diversas áreas. O termo IA, criado por John McCarthy (1956), ganhou várias definições ao longo dos anos, mas pode-se considerar que é "o estudo de como fazer com que os computadores façam coisas que, no momento, as pessoas fazem melhor" (Rich; Knight, 1991). A área da IA deu seus primeiros passos com o

8 *Article 29 Working Party*: O Grupo de Trabalho do Artigo 29 foi criado pela Diretiva nº 95/46/CE e tratou de questões relacionadas à proteção da privacidade e dos dados pessoais até 25 de maio de 2018, data da entrada em vigor do RGPD. Disponível em: https://edpb.europa.eu/about-edpb/more-aboutedpb/article-29-working-party_en.

desenvolvimento do primeiro modelo de um neurônio artificial (McCulloch; Pitts, 1943), a criação da primeira máquina para simular uma rede neural artificial (Minsky, 1952) e os avanços na área de aprendizado, estendendo o conceito de neurônios artificiais (Rosenblatt, 1961).

Apesar de ser uma área relativamente antiga, atraiu os holofotes da mídia e da sociedade somente nas últimas décadas com a IBM e seu Deep Blue vencendo o campeão de xadrez Garry Kasparov (1997) e o Watson competindo e vencendo o jogo "Jeopardy!" (2011), com os algoritmos de recomendações da empresa Amazon (Linden, 2003), com as pesquisas do Google na área de reconhecimento de objetos em imagens (Quoc, 2012), e com o DeepMind propondo soluções para um problema que estava aberto por mais de 50 anos na área de enovelamento de proteínas (Jumper *et al.*, 2021) e na área de previsão do tempo em curto prazo (Ravuri, 2021).

Contudo, quando sistemas de IA são utilizados para fins de avaliações educacionais, campanhas políticas, seleção de candidatos a empregos, e até mesmo para a implementação de mecanismos relacionados a policiamento preditivo e a admissão de migrantes em um Estado, diversas considerações devem ser feitas. Uma delas diz respeito à falta de transparência em relação aos titulares dos dados ou destinatários desses sistemas, que, muitas vezes, não conseguem acessar ou avaliar a qualidade das inferências que são realizadas sobre eles. Tal situação coloca em risco questões fundamentais inerentes à dignidade humana, como liberdade, igualdade, não discriminação, devido processo legal, proteção dos dados pessoais e dados pessoais sensíveis e da autodeterminação informacional.

Para Laurianne-Marie Schippers e Marília Papaléo Gagliardi, muitas vezes existe a concepção de que algoritmos seriam imparciais e, portanto, seriam ideais para tomar decisões no lugar de seres humanos, que poderiam apresentar decisões enviesadas e prejudicadas por suas próprias percepções. Se é uma máquina que está fazendo sugestões de conteúdo, ou organizando o sistema de identificação, esta deveria, em tese, não possuir nenhum conteúdo discriminatório em sua origem (Schippers; Gagliardi, 2021, s.p.).

No entendimento de Ana Frazão, a transferência ou delegação total do processo decisório de agentes públicos e privados para sistemas algorítmicos é procedimento que envolve diversos riscos, considerando as limitações já apontadas na programação e nos *designs* de tais sistemas. Para a autora, mesmo quando não ocorre a terceirização total e o sistema algorítmico tem o papel de ser apenas um auxiliar no processo decisório, permanecen-

do o ser humano com a "última palavra", os desafios não são banais. Afinal de contas, pouco se sabe sobre como se comportam seres humanos diante de decisões algorítmicas, havendo o fundado receio de que tendam a concordar com os seus resultados, até porque, considerando a opacidade dos algoritmos, não conseguem compreendê-los nem os questionar (2021, s.p.).

De acordo com Ana Frazão, o chamado fenômeno do viés da automação (*automation bias*) sugere que

> [...] ferramentas de automação influenciam decisões humanas de formas significativas e geralmente ruins. Dois tipos de erros são particularmente comuns: (i) erros de omissão, nos quais as pessoas não reconhecem quando os sistemas automatizados erram e (ii) erros comissivos, nos quais as pessoas seguem os sistemas automatizados sem considerar suas informações contraditórias. (2021, s.p.)
>
> Consequentemente, a autora entende que uma forte confiança em sistemas automatizados pode "alterar as relações das pessoas com as suas tarefas, criando uma espécie de 'para-choque' entre as decisões e os seus impactos, com a consequente perda do senso de responsabilidade e da *accountability*" (2021, s.p.).

Neste contexto, Zygmunt Bauman reflete sobre a adiaforização, em que sistemas e processos se divorciam de qualquer consideração de caráter moral. Para o autor, um ângulo da adiaforização em termos de vigilância é a forma como dados do corpo (dados biométricos, DNA) ou por ele desencadeados (por exemplo, situações em que faz um *login*, usa-se um cartão de acesso ou mostra-se a identidade) são sugados para bases de dados a fim de serem processados, analisados, concatenados com outros dados e depois cuspidos de volta como "replicação de dados" (2013, p. 15).

As informações que fazem as vezes da pessoa são constituídas de "dados pessoais" apenas no sentido de que se originaram em seu corpo e podem afetar suas oportunidades e escolhas existenciais. A "replicação e fragmentação de dados" tende a inspirar mais confiança que a própria pessoa – que prefere contar sua própria história. Os *designers* de *software* dizem que estão simplesmente "lidando com dados", de modo que seu papel é "moralmente neutro" e suas avaliações são apenas "racionais" (2013, p. 15).

De uma forma geral, o uso de tecnologias de decisões automatizadas pode incrementar a celeridade e a acurácia na análise de uma quantidade

expressiva de processos e situações, contemplando princípios relevantes de celeridade, eficiência e economia. Entretanto, acurácia e celeridade não podem ser as métricas-fim ou fundamento raiz da utilização de sistemas de inteligência artificial, mas deve existir uma associação com a sustentabilidade, a inclusão e proteção da diversidade, a solidariedade e a equidade (Peixoto, 2020, p. 318).

Diante do exposto, percebe-se que os dados pessoais, a vigilância e as decisões automatizadas são temas que possuem grandes conexões entre si, e que podem ser vistos em diversos segmentos da sociedade atual. Nesse contexto, imperioso lembrar os ensinamentos de Shoshana Zuboff sobre o capitalismo de vigilância e como ele é visto como um movimento que visa impor uma nova ordem coletiva baseada em uma certeza total. A autora informa que somos "as fontes do superávit crucial do capitalismo de vigilância: os objetos de uma operação de extração de matéria-prima tecnologicamente avançada e da qual é cada vez mais impossível escapar" (2020, p. 22). Sob essa perspectiva, o próximo tópico busca compreender como as técnicas de perfilamento são utilizadas no ambiente do controle migratório e como elas podem impactar os migrantes.

O USO DE TECNOLOGIAS DE PERFILAMENTO NO ÂMBITO DO CONTROLE MIGRATÓRIO

Nesta perspectiva, a vigilância é uma dimensão-chave do mundo moderno. Conforme Zygmunt Bauman, viajantes em passagem por aeroportos sabem que precisam atravessar não apenas o controle de passaportes em sua versão do século XXI, mas também por novos dispositivos, como escâneres corporais e aparelhos de checagem biométrica, que têm proliferação desde os atentados de 11 de setembro nos Estados Unidos (2013, p. 9).

Neste contexto, a segurança transformou-se em um empreendimento orientado para o futuro e funciona por meio da vigilância, tentando monitorar o que *vai* acontecer pelo emprego de técnicas digitais e raciocínio estatístico. A segurança funciona a distância tanto no espaço quanto no tempo, circulando de maneira fluida, juntamente com os Estados-Nação, mas para além deles, em um domínio globalizado. Na visão de David Lyon, "processos de estereotipia e medidas de exclusão estão à espera dos grupos desafortunados o bastante para serem rotulados de 'indesejados'" (2013, p. 13).

As dimensões tecnológicas, ou melhor, tecnossociais, da vigilância atual relacionam-se com a classificação da população em categorias, tendo em vista um tratamento diferencial. Surgem questões relacionadas a como

as tecnologias provocam consequências menos catastróficas, porém não menos insidiosas, em particular para grupos já discriminados.

Alguns governos ao redor do mundo já se valem de ferramentas de inteligência artificial para auxiliar na tomada de decisões para realizar controles migratórios. Neste sentido, o texto observa as atividades dos governos do Reino Unido e do Canadá a fim de identificar como essas tecnologias de perfilamento estão sendo utilizadas e quais os potenciais impactos aos direitos humanos dos migrantes. Por fim, observaremos as iniciativas no Brasil, com destaque ao projeto Embarque Mais Seguro.

No Reino Unido, a triagem e definição da fila de pedidos de visto era realizada de forma automatizada, por meio de um sistema que classificava os pedidos, a partir das informações fornecidas pelos aplicantes, em três possíveis bandeiras: vermelha, amarela e verde. De acordo com o Home Office, órgão do Reino Unido responsável pela imigração, dentre outras atribuições relacionadas à segurança, a análise realizada pelo sistema servia como um auxílio às decisões de concessão de visto a serem tomadas pelos agentes. Para Henry McDonald, do *The Guardian*, a decisão final era tomada por pessoas naturais, com base na separação e sistematização feita pelo algoritmo (BBC, 2020, s.p.).

Em 2020, diversas notícias surgiram alegando que o algoritmo utilizado estaria fazendo classificações negativas e de cunho racista. Conforme a BBC, a acusação formada pelo Joint Council for the Welfare of Immigrants[9] e o grupo Foxglove[10] era a de que a categorização realizada pelo sistema levaria em conta, dentre outras informações, a nacionalidade do migrante. Nesse sentido, o próprio sistema de migração do Reino Unido (Home Office) utilizaria uma lista de nacionalidades suspeitas que seriam automaticamente classificadas com a bandeira vermelha (BBC, 2020, s.p.).

A apuração partiria de uma premissa de dificultar o processo de obtenção de vistos de imigrantes com base na nacionalidade, e acarretaria

9 O Conselho Conjunto para o Bem-Estar dos Imigrantes (JCWI) é uma instituição de caridade independente do Reino Unido que fornece informações sobre imigração e aconselhamento jurídico gratuito sobre questões relacionadas à imigração, incluindo direito a benefícios (Barnet, 2021).

10 Foxglove é uma organização independente e sem fins lucrativos do Reino Unido. Em seus primeiros dois anos, construiu um histórico de responsabilização de grandes empresas de tecnologias e governos. A organização forçou a divulgação de contratos secretos entre gigantes da tecnologia e o governo, conhecidos como "NHS Covid-19 data deals", parou o algoritmo racista de vistos do Home Office e ajudou a tornar a avaliação justa para todos os alunos no Reino Unido, desafiando o algoritmo de Ofqual. Foxglove é formada por uma equipe de advogados especialistas em tecnologia e especialistas em comunicação (Foxglove, 2021).

problemas como, por exemplo, as análises de concessão de vistos serem muito mais demoradas, passarem por uma averiguação mais severa e terem uma chance maior de serem negadas (BBC, 2020, s.p.). De acordo com o Foxglove, essa discussão tem ocorrido desde 2017 no sistema de justiça britânico. Todavia, apenas em agosto de 2020, com o processo ainda em curso, o departamento do Home Office optou por parar o uso do algoritmo, a fim de auditá-lo para identificar a presença de potenciais vieses que gerassem discriminação e qualquer tomada de decisão negativa de forma injustificada (Heaven, 2020, s.p.).

Já no contexto do Canadá, o The Citizen Lab, em conjunto com o International Human Rights Program da Faculdade de Direito da Universidade de Toronto, publicou uma pesquisa de Petra Molnar e Lex Gill chamada *Bots at the gate: a Human Rights analysis of automated decision-making in Canada's immigration and refugee system*. O relatório elaborado em 2018 indica que o governo canadense também se utiliza de ferramentas de tomadas de decisão automatizada no âmbito migratório e de refúgio (Molnar; Gill, 2018, p. 14).

O documento aponta para a possibilidade e riscos de usos de ferramentas de tomadas de decisão automatizadas em diversos momentos do fluxo migratório. Entretanto, há a menção de que o governo emprega tais ferramentas na análise dos solicitantes de visto, mas não é possível verificar como essa análise é feita e quais são os critérios utilizados (2018, p. 24-5). Sob essa perspectiva, é possível afirmar que há falta de transparência nos procedimentos, impossibilitando, inclusive, que a sua neutralidade seja verificada.

O relatório também destaca que o algoritmo pode ser utilizado para fazer classificações sobre a veracidade do alegado por um candidato a visto. Por exemplo, se ele realmente é casado (2018, p. 25-33). Essa decisão influencia diretamente na credibilidade do aplicante, e pode ser imprescindível para a concessão ou não do visto.

O documento não se posicionou de forma repulsiva à própria tecnologia, mas ao uso irresponsável dela, podendo exacerbar disparidades. Identificou-se que o governo canadense, desde 2014, ampliou o uso de tecnologia e está desenvolvendo um sistema de análise preditiva para automatizar atividades até então conduzidas por funcionários de imigração (2018, s.p.). A pesquisa afirma que o Canadá tem obrigações domésticas e internacionais claras com o respeito e a proteção dos direitos humanos e

que cabe aos políticos, funcionários públicos, tecnólogos e engenheiros, assim como advogados, sociedade civil e universidade, adotarem uma ampla visão crítica dos impactos reais do uso de tecnologia sobre a vida humana (2018, s.p.). Para os autores, o desafio não é como usar novas tecnologias para consolidar velhos problemas, mas, em vez disso, para melhorar entender como podemos usar esta oportunidade para imaginar e projetar sistemas mais transparentes, equitativos e justos (2018, p. 7).

Na visão de Fabiano Hartmann Peixoto, o tema apresenta fundamento jurídico, já que as várias facetas do uso de sistemas de decisão automatizados podem atingir direitos humanos, incluindo direitos à igualdade e à não discriminação; liberdade de movimento, expressão, religião e associação; privacidade, vida, liberdade e segurança das pessoas. Para o autor, a temática também desperta questões de direito constitucional e administrativo, acesso à justiça, responsabilidade público e privada, capacidade de gestão pública e governamental, bem como outros impactos globais (2020, p. 309).

Diante do exposto, percebe-se que as tecnologias de perfilamento envolvendo TRF e IA já estão sendo utilizadas no ambiente do controle migratório em diversos lugares do mundo[11]. Nesse contexto, cabe frisar que os exemplos citados são casos chave para observar o desenvolvimento da temática; entretanto, não são os únicos: Estados Unidos, Nova Zelândia, Austrália, entre outros países, também utilizam tecnologias similares. Na sequência, observa-se como a temática se desenvolve no Brasil.

INICIATIVAS EM SOLO NACIONAL

Inicialmente, destaca-se o projeto piloto Embarque Mais Seguro. Trata-se de um sistema de reconhecimento por biometria, que valida a identidade do viajante por *selfies* tiradas na hora comparadas com bases de dados públicas preexistentes do Denatran e do Barramento SGD (TSE)[12]. Conforme informações divulgadas pela Serpro, empresa de inteligência em tecnologia

11 Saiba mais em: SCHIPPERS, Laurianne-Marie; GAGLIARDI, Marília Papaléo. *Inteligência artificial e controle migratório*: algoritmos podem discriminar migrantes? Publicado em: 16 jan. 2021. Disponível em: https://medium.com/o-centro-de-ensino-e-pesquisa-em-inova%C3%A7%C3%A3o-est%C3%A1/intelig%-C3%AAncia-artificial-e-controle-migrat%C3%B3rio-algoritmos-podem-discriminar-migrantes-85d04-d152440. Acesso em: 10 set. 2021.

12 O barramento de serviços é uma solução centralizada que permite que um órgão envie processos ou documentos administrativos digitais para outro de maneira segura e com confiabilidade de entrega. Tal solução permite o trâmite eletrônico entre plataformas distintas (SEI, 2021).

da informação (TI) do governo federal, em parceria com o Ministério da Infraestrutura (MInfra), o objetivo do projeto é tornar o processo de embarque nos aeroportos mais eficiente e as viagens aéreas mais seguras. Para garantir os resultados, estão sendo realizados projetos piloto em vários aeroportos do País. No momento, o projeto já foi implementado nos aeroportos de Florianópolis (SC), Salvador (BA), Belo Horizonte (Confins), Santos Dumont (RJ) e Congonhas (SP) (Serpro, 2021, s.p.).

A finalidade do serviço é a segurança da informação, a prevenção de fraudes, o aumento da proteção à integridade física no transporte aéreo, a facilitação do transporte aéreo, o aumento da eficiência e da velocidade dos processos aeroportuários. Além disso, os dados coletados durante o piloto serão utilizados para estudos estatísticos visando à melhoria do serviço (2021, s.p.).

O projeto encontra-se alinhado com as principais iniciativas e projetos internacionais do setor, como Programa de Identificação de Viajantes (*Traveller Identification Programme – TRIP*) da Organização da Aviação Civil Internacional (OACI) e *One ID* da Associação Internacional do Transporte Aéreo (IATA) (2021, s.p.).

De acordo com a Serpro, o uso de todas as informações coletadas está alinhado à Lei Geral de Proteção de Dados Pessoais (LGPD). No *site* do projeto, é possível identificar a área "Aviso de Privacidade", em que encontram o nome e o endereço do controlador, o nome e o endereço do encarregado de dados, informações gerais sobre o tratamento dos dados, dados de oferta do serviço, geração de dados de *login* e direito dos titulares de dados (2021, s.p.). Nesse contexto, a empresa afirma que, em relação ao tratamento de dados, atua em consonância com

> [...] sua missão institucional, respeitando o direito fundamental à privacidade e visando o melhor uso da tecnologia da informação para a satisfação da sociedade e de seus clientes, e a sustentabilidade e autonomia empresarial, garantindo a segurança, estabilidade e a continuidade de seus serviços. (2021, s.p.)

A Serpro afirma que busca garantir que os dados pessoais sejam tratados sempre em conformidade com as bases legais da LGPD e garante que os dados são tratados estritamente para as finalidades informadas, sem desvio na direção de outros propósitos. Com relação às bases legais para o tratamento de dados, a empresa reconhece como bases legais o cumprimento de obrigação legal, a execução de políticas públicas, o legítimo interesse ou a

garantia da prevenção à fraude e à segurança do titular, dentre as hipóteses previstas nos arts. 7º e 11 da LGPD (2021, s.p.).

Quando o assunto é a segurança da informação e privacidade, a empresa informa que adota controles e procedimentos de segurança de forma a assegurar a confidencialidade, integridade, disponibilidade e privacidade dos dados sob sua responsabilidade. (2021, s.p.).

Em relação ao compartilhamento de dados coletados, a Serpro informa que compartilha com operadores e terceiros. As empresas fornecedoras de equipamentos biométricos são consideradas operadores de dados pessoais, pois tratam dados em nome do controlador. Os dados pessoais compartilhados com estes operadores são apenas aqueles necessários para o funcionamento dos equipamentos biométricos no contexto do tratamento (foto frontal, com data/hora e sigla do aeroporto onde foi coletada; dados do cartão de embarque: número do voo, nome, data, trecho), na medida em que apenas capturam tais dados, transmitem e recebem uma resposta de ação a ser executada. Nem todos os dados mencionados são tratados por todos os operadores indicados. A finalidade do compartilhamento com tais operadores guarda relação com o próprio objetivo do serviço Embarque Mais Seguro, já que, para a viabilização deste, é necessária a participação de fornecedores especializados neste tipo de equipamento (2021, s.p.).

Em relação aos terceiros, a empresa informa que os dados tratados durante o processo são compartilhados com a Secretaria de Aviação Civil (SAC/MInfra), que possui a finalidade pública de administrar a aviação civil no território nacional e executar, dentro de suas competências legais, estudos e análises estatísticas baseadas nos dados coletados no projeto em questão. A base legal é a prevista no art. 7º, III, da LGPD, definida nas situações de tratamento e uso compartilhado de dados necessários à execução de políticas públicas (2021, s.p.).

A Serpro informa que todos os agentes de tratamentos e terceiros participantes do projeto firmam termo de tratamento de dados pessoais em que são estipulados os deveres e as responsabilidades dos participantes. Apesar de a página da Serpro indicar que o projeto está alinhado à Lei Geral de Proteção de Dados, na prática não é exatamente o que ocorre (2021, s.p.).

Consoante informações do Instituto de Referência em Internet e Sociedade (IRIS), pesam sobre o projeto Embarque Mais Seguro vários questionamentos relativos ao tratamento dos dados pessoais. Tais questionamentos incluem a legitimidade do compartilhamento de dados entre o Denatran

e o Serpro, até a ausência de esclarecimentos sobre a tecnologia utilizada no Programa e, por fim, informações relativas e elaboração do Relatório de Impacto à Proteção de Dados Pessoais (RIPD) (IRIS, 2020, s.p.).

Neste contexto, serão abordadas brevemente algumas lacunas nas informações da Serpro e do MInfra sobre a tecnologia de reconhecimento facial em aeroportos e o projeto Embarque Mais Seguro. De acordo com o IRIS, a elaboração do RIPD deve ocorrer antes do tratamento de dados pessoais dos titulares. O Instituto questiona se

> [...] foi feito um relatório de impacto pelo Ministério da Infraestrutura e Serpro antes de ser feito qualquer teste com pessoas naturais no Aeroporto Internacional de Florianópolis? Se sim, por que o relatório de impacto não foi publicizado ou por que ele não foi sequer mencionado? Se não, por que o relatório de impacto não foi elaborado? (2020, s.p.)

Ambos, Serpro e MInfra, estão sujeitos às obrigações da LGPD como agentes de tratamento de dados pessoais. Sobre o tema, o IRIS lembra o caso do IBGE de compartilhamento de dados, que foi julgado pelo Supremo Tribunal Federal (STF) em 2020 – na ADIn 6387/2020. O caso consolidou o entendimento de que, em situações que irão envolver o tratamento de um grande volume de dados pessoais da população brasileira, é fundamental que um Relatório de Impacto seja elaborado antes de qualquer operação de tratamento (2020, s.p.). No caso do Embarque Mais Seguro, já houve o compartilhamento da base de dados do Denatran com a Serpro e não é possível, inclusive, saber se foi realizado um relatório de impacto para avaliar os riscos associados a esse compartilhamento (2020, s.p.).

Na visão do Laboratório de Políticas Públicas e Internet (LAPIN), o emprego irrestrito de tecnologias de vigilância impacta negativamente os direitos fundamentais dos indivíduos. Isso porque, além dos impactos diretos sobre a privacidade, o tratamento de dados advindo do uso de tais tecnologias tende a refletir vieses algoritmos que reforçam discriminações e a impactar o direito à proteção de dados da população. As avaliações de risco, portanto, são instrumentos a serem utilizados pelos agentes de tratamento para discernir sobre a licitude da atividade e compreender os riscos que se impõem (Lapin, 2021, p. 31).

Outro ponto de questionamento é por que a base de dados do Denatran foi compartilhada com a Serpro e está sendo usada para verificação de identidade dos passageiros nos aeroportos por meio de tecnologias de reconhecimento facial. Consoante o Instituto, a base de dados do

Denatran possui atualmente dados pessoais de cerca de 78 milhões de pessoas. Essa mesma base de dados foi alvo de controvérsias anteriores, como, por exemplo, o vazamento de dados em 2019 no Detran RN e no Detran RJ, além da solicitação pela Agência Brasileira de Inteligência (ABIN) à Serpro de todas as CNHs no Brasil (2020, s.p.).

Por fim, o IRIS entende que o País precisa, urgentemente, mudar a mentalidade de desenho e implementação de políticas públicas que se valem de tecnologias como as de reconhecimento facial, a fim de que a garantia dos direitos possa ser preservada e avanços na prestação de serviços públicos possam ser feitos, desde que sempre resguardados os direitos de todas as pessoas envolvidas (2020, s.p.). Conforme o Relatório do Instituto, a postura que vem sendo adotada é sempre "compartilhar bases de dados, implementar tecnologias de reconhecimento facial e depois pensar em talvez fazer um relatório de impacto, e isso precisa urgentemente mudar" (2020, s.p.).

No Brasil, outro uso comum de TRF em aeroportos é o da Receita Federal (RFB). Desde 2016, a Receita utiliza o sistema de reconhecimento facial da empresa NEC em 14 aeroportos internacionais, como os de Brasília, Guarulhos, Recife, Rio de Janeiro e Salvador (NEC, 2016, s.p.). A contratação ocorreu via licitação, pela modalidade de pregão eletrônico, e o valor do contrato foi de R$ 7.576.090,72 (Brasil, 2015, s.p.).

O objeto do contrato foi o fornecimento de solução de TRF para localizar viajantes com risco aduaneiro identificado. Dessa forma, "os servidores dos órgãos de controle podem identificar inequivocamente os indivíduos de potencial interesse (previamente selecionados pelo sistema de gerenciamento de risco) e encaminhá-los para fiscalização minuciosa" (Lapin, 2021, p. 61).

Para tanto, a Receita informou que

> a solução tecnológica se processa exclusivamente de forma interna no ambiente computacional da RFB, que é acessado apenas por servidores públicos cadastrados com certificado digital. A RFB não compartilha os dados da solução de reconhecimento facial nem com a iniciativa privada nem com outro órgão público.

Ainda, a Receita afirmou que não existem mecanismos no sistema para extrair os dados coletados mediante solicitação do titular (Lapin, 2021, p. 61).

Há, ainda, o projeto piloto de Cidades Inteligentes, que inclui sistema inteligente de controle, monitoramento e segurança, chamado Fronteira Tech, inaugurado em dezembro de 2019, em parceria com a Receita Federal e o Instituto de Desenvolvimento Tecnológico. O projeto funciona de forma integrada com o banco de dados da Receita Federal e reforça o controle aduaneiro na Ponte da Amizade, em Foz do Iguaçu (PR), na fronteira entre Brasil e Paraguai, e utiliza:

– 35 luminárias inteligentes, com duas câmeras em cada, totalizando 70 equipamentos com capacidade de fazer reconhecimento facial e identificar placas de automóveis;

– 4 câmeras fixas com as mesmas tecnologias instaladas em pontos estratégicos;

– 15 luminárias de LED com telegestão e GPS;

– 11 sensores de tiro;

– *software* de inteligência artificial que identifica padrões e gera dados que ajudam no combate ao contrabando e ao tráfico de drogas e armas;

– identifica e alerta para a placa de um veículo roubado e a identificação facial para procurados da justiça (ABDI, 2021a, s.p.).

Em setembro de 2021, o projeto foi implementado nas áreas de fronteira de Pacaraima, em Roraima (RR), na divisa do Brasil com a Venezuela. O investimento da Agência Brasileira de Desenvolvimento Industrial (ABDI) na região de Pacaraima é da ordem de R$ 3,1 milhões. Os recursos são destinados à instalação e manutenção dos equipamentos, que incluem luminárias inteligentes com dimerização (10); luminárias inteligentes com câmeras de vigilância integradas (20); *software* de reconhecimento facial; câmeras de sensoriamento do tipo *speed dome* (4); *datacenter* para armazenamento e processamento de imagens e dados; telas de *videowall*; câmeras de reconhecimento de placas de veículos (4); *software* de reconhecimento de placas; drone com câmera termográfica, além da licença dos *softwares* por três anos (ABDI, 2021b, s.p.).

Nos casos acima mencionados, portanto, é possível notar a utilização e a implementação de algoritmos em diferentes momentos do processo migratório. Uma das finalidades observadas seria aquela de acelerar e/ou facilitar os processos migratórios; contudo, há risco com respeito à possi-

bilidade de erro do sistema[13]. Identificar uma pessoa como sendo outra ou simplesmente não a identificar pode levar a situações de discriminação ou restrição injustificada de direitos. O reconhecimento facial corre o risco de ser transformado em arma pela aplicação da lei contra comunidades marginalizadas em todo o mundo. De Nova Delhi a Nova York – passando pelo Brasil –, essa tecnologia invasiva poderia "virar nossas identidades contra nós, além de minar os direitos humanos" (Mahmoudi, 2021). Ou seja, perfilar alguém erroneamente em um contexto do controle migratório, por exemplo, pode levar a abordagens e apreensões indevidas, além das violações de direitos humanos[14], e casos de racismo[15], transfobia e islamofobia.

Diante do exposto, compreende-se que o uso de tecnologias de perfilamento no ambiente do controle migratório pode auxiliar na celeridade do processo; entretanto, o uso de sistemas automatizados precisa seguir diretrizes que respeitem os direitos humanos dos migrantes, com ênfase na não discriminação, no tratamento igualitário, na transparência das decisões e na dignidade da pessoa humana.

O TRATAMENTO DOS DADOS PESSOAIS NO CONTEXTO MIGRATÓRIO BRASILEIRO

Os incrementos tecnológicos dos últimos anos criaram modelos sofisticados de tratamentos de dados pessoais. Vivemos em uma sociedade orientada por dados, conhecida como *data-driven society*, em que o uso de serviços de empresas, no campo da comunicação, do comércio, do turismo ou entretenimento, sujeita-se ao tratamento de dados. Garantir a proteção de dados do indivíduo é fundamental nesse contexto (Viola; Heringer; Carvalho, 2021, p. 4).

Na visão de Samuel Oliveira, nos últimos anos também se intensificaram as discussões sobre a questão da segurança. Nesse cenário, ganhou

13 ROGERS, Lindsay. *Facial recognition technology at airports isn't even working*. Publicado em: 15 fev. 2021. Disponível em: https://www.insidehook.com/daily_brief/travel/airport-facial-recognition-not-working. Acesso em: 6 dez. 2021.

14 No documentário *Coded Bias*, disponível na plataforma de *streaming* Netflix, a pesquisadora Joy Buolamwini apresenta um relato técnico e ao mesmo tempo uma denúncia social acerca dos vieses presentes em algoritmos de reconhecimento facial.

15 *Conforme Bruno Souza, a maioria das câmeras que fazem reconhecimento facial na cidade do Rio de Janeiro foram instaladas no bairro de Copacabana – o cartão-postal da cidade. No segundo dia do uso, uma mulher foi presa ao ser identificada erroneamente pelo sistema, que apontou mais de 70% de semelhança entre ela e Maria Leda, uma pessoa foragida da Justiça. Entretanto, a verdadeira criminosa estava presa desde 2015. As polícias militar e civil do Rio de Janeiro utilizaram um banco de dados desatualizado. Esse caso é emblemático porque expõe a falha da máquina de leitura biométrica facial e a irresponsabilidade por parte da secretaria de segurança pública (Souza, 2021).*

força a ideia de que um afrouxamento na proteção de dados pessoais seria uma maneira eficaz de se combater a violência e até mesmo o terrorismo (Oliveira, 2021, p. 133). Para Stefano Rodotà, se seguirmos esse raciocínio, "a questão corre o risco de ser posta de maneira imprópria, como se segurança e proteção de dados fossem valores incompatíveis e como se a tutela de um excluísse automaticamente qualquer relevância do outro" (Rodotà, 2004, p. 95).

A tutela do direito à proteção de dados e à privacidade dos migrantes explicita uma questão latente na regulação jurídica do tema, que é a da composição e tensão entre os direitos do titular dos dados e o interesse público ou estatal (Gediel; Corrêa, 2021 p. 619). Diante do exponencial progresso do uso de sistemas automatizados de decisão no ambiente do controle migratório, surgem preocupações quanto aos seus riscos e aos meios regulatórios adequados. Nesse cenário, faz-se necessário destacar que existem os seguintes tratamentos de dados no contexto migratório: (i) o tratamento de dados pessoais realizado por companhias aéreas para a execução dos seus serviços; (ii) o tratamento com o objetivo da segurança pública[16]; e também (iii) o tratamento voltado para a segurança nacional[17].

Inicialmente, cabe lembrar que o arcabouço legislativo nacional já conta com normativas que versam sobre a proteção da privacidade, seja na esfera constitucional, seja na infraconstitucional. Nesta senda, destacam-se a Constituição Federal de 1988 (que apontou a inviolabilidade da vida privada, o sigilo das comunicações, além do *habeas data* como instrumento

16 A segurança pública tem um capítulo próprio na Constituição Federal de 1988, que está contido no Título V, "Da Defesa do Estado e das Instituições Democráticas". O capítulo III do Livro V, "Da Segurança Pública", consigna somente o art. 144, onde se extrai a definição constitucional do conceito de segurança pública, explícita no *caput*: "A segurança pública, dever do Estado, direito e responsabilidade de todos, é exercida para a preservação da ordem pública e da incolumidade das pessoas e do patrimônio" (Brasil, 1988). Ao atribuir ao Estado o dever pela segurança pública, reconhece-o como serviço público a ser garantido pela máquina estatal, direito inalienável de todos os cidadãos. Já a definição da segurança também como responsabilidade de todos pode ser interpretada à luz da necessidade de que haja participação social nas políticas públicas relacionadas a esse campo. Adicionalmente, é possível compreender que a segurança pública não pode ser vista apenas como atribuição do Estado, uma vez que a sociedade tem um papel importante não somente na participação e no controle das políticas, como também na socialização dos indivíduos, na perpetuação dos mecanismos informais de controle social e de autocontrole, a partir da perspectiva de que não é somente o controle pelo Estado que garante a segurança de todos (Fontoura; Rivero; Rodrigues, 2015, p. 143).

17 Segurança nacional não se confunde com segurança pública. Entende-se aqui segurança nacional como um estado em que se percebe, materialmente: a) a estabilidade e a inviolabilidade dos limites fronteiriços do Estado; b) a capacidade de se traduzir a soberania nacional, bem como a capacidade nacional de projetar poder no exterior, em um conjunto de medidas que proporcione ganhos sociais e econômicos para a população nacional; c) a solidez e a impessoalidade do sistema constitucional, assim como sua impermeabilidade em relação a pressões externas; e d) a garantia da previsibilidade legal das relações político-eleitorais e econômicas (Costa, 2018, p. 124-125).

apto a assegurar a proteção de informações e dados pessoais), o Código Civil de 2002 (que protege diretamente a vida privada), o Código de Defesa do Consumidor de 1990 (que o faz, dedicando a Seção VI à proteção de bancos de dados e de cadastros dos consumidores), além do Marco Civil da Internet de 2014, que trouxe dispositivos destinados à proteção da privacidade, que, por sua vez, constitui um dos pilares da lei.

A própria Lei de Migração (Lei nº 13.445/2017) faz referência expressa ao direito de proteção de dados pessoais do migrante e lhe assegura, em seu art. 4º, XIII, direito à informação e à confidencialidade, com remissão ao disposto na Lei de Acesso à Informação (LAI – Lei nº 12.527/2011). A LAI estabeleceu as regras para o tratamento dos dados pessoais pelo Poder Público, que fica submetido ao dever de transparência e de respeito à "[...] à intimidade, vida privada, honra e imagem das pessoas, bem como às liberdades e garantias individuais" (art. 31 da Lei nº 12.527/2011). A tutela de tais direitos deve ser pensada na sua vinculação aos princípios da Lei de Migração, previsto em seu art. 3º, sobretudo na vedação à xenofobia, à discriminação e à criminalização da migração (Brasil, 2017; Brasil, 2011).

Além disso, está em tramitação, no Congresso Nacional, o Projeto de Emenda Constitucional (PEC) nº 17/2019, aprovado pela Câmara e em tramitação no Senado e que visa incluir a proteção de dados pessoais entre os direitos e garantias fundamentais, além de fixar a competência privativa da União para legislar sobre a proteção e tratamento de dados pessoais (Senado, 2021, s.p.).

Se o País já conta com tantas legislações sobre o tema, por que criar mais uma? Na visão de Eduardo Magrani, a regulação existente é insuficiente para "proteger os dados pessoais e a privacidade em suas mais diversas facetas. Daí a importância da LGPD, que veio preencher as lacunas da legislação e é aplicável a uma gama mais ampla de usos da internet" (2019, p. 87).

Nesta senda, pontua-se que a Lei nº 13.709, de 14 de agosto de 2018, conhecida como Lei Geral de Proteção de Dados (LGPD), foi influenciada por legislações semelhantes adotadas nos Estados Unidos da América (EUA) – a *California Consumer Privacy Act* (CCPA) – e na União Europeia, denominada *General Data Protection Regulation* (GDPR) e segue a tendência mundial do aumento do foco em privacidade e proteção de dados.

Na visão do Instituto de Tecnologia e Sociedade do Rio de Janeiro (ITS Rio), a LGPD trouxe um novo paradigma que passou a fundamentar a

abordagem do direito à privacidade, centrado no ideal de autodeterminação informativa, autonomia e controle do cidadão de seus dados. Esse novo ambiente regulatório busca harmonizar a proteção dos direitos dos indivíduos e a provisão de segurança jurídica nas relações permeadas pelo tratamento de dados pessoais, em um mundo hiperconectado e marcado pelo vigilantismo (ITS, 2021, s.p.).

O primeiro cenário (o tratamento de dados pessoais realizados por companhias aéreas) é contemplado pela LGPD, considerando que a lei dispõe sobre o

> tratamento de dados pessoais, inclusive nos meios digitais, [...] *por pessoa jurídica de direito público ou privado*, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. (Brasil, 2018, s.p. – grifos nossos)

O segundo e o terceiro cenários, por sua vez, não encontram proteção na LGPD. Neste cenário, é preciso pontuar que, no momento da construção da lei, os legisladores preferiram deixar de fora o tratamento de dados para fins de segurança pública, defesa nacional, segurança do Estado ou atividades de investigação e repressão de infrações penais, conforme disposto no art. 4º, III, da LGPD.

O § 1º do artigo em comento afirma que o tratamento de dados pessoais "previsto no inciso III será regido por legislação específica, que deverá prever medidas proporcionais e estritamente necessárias ao atendimento do interesse público, observados o devido processo legal, os princípios gerais de proteção e os direitos do titular previstos nesta lei" (Brasil, 2018, s.p.).

As peculiaridades dessa lei específica são justificadas pelo desafio de se garantir um equilíbrio entre a investigação penal, atividade que demanda tratamento de dados de diversos atores, e os direitos fundamentais de privacidade e proteção de dados (Costa; Reis, 2021, s.p.). Nesse contexto, é preciso compreender que não há dados pessoais insignificantes e que há

> um direito autônomo à proteção de dados pessoais e o seu duplo efeito sobre os deveres do Estado (um dever negativo de não interferir indevidamente no direito fundamental e um dever positivo de adotar medidas positivas para a proteção desse direito). (Mendes, 2020, s.p.)

Em novembro de 2020, um anteprojeto de lei sobre a temática foi apresentado à Presidência da Câmara dos Deputados, conhecido como Anteprojeto da LGPD Penal. O anteprojeto possui doze capítulos e 68

artigos, divididos em oito eixos temáticos: (i) âmbito de aplicação da lei; (ii) condições de aplicação; (iii) base principiológica; (iv) direitos e obrigações; (v) segurança da informação; (vi) tecnologias de monitoramento; (vii) transferência internacional de dados; e (viii) a autoridade de supervisão.

O texto teve duas inspirações principais: a LGPD e a Diretiva nº 2016/680 da União Europeia, cujo objeto é similar àquele do anteprojeto. Especificamente para o capítulo VII, sobre tecnologias de monitoramento, a inspiração veio de leis estadunidenses da cidade de Nova York e do estado de Washington (Costa; Reis, 2021, s.p.).

O anteprojeto regula somente as alíneas *a* e *d* (segurança nacional e atividades de investigação e repressão de infrações penais, respectivamente), deixando para regulação posterior as alíneas *b* e *c*, referentes aos tratamentos de dados para defesa nacional e segurança do Estado. Essa lacuna demanda a criação de regulação o mais brevemente possível, a fim de proteger, de forma adequada, os direitos dos titulares de dados individual e coletivamente.

Na visão de Laura Schertel Mendes, relatora da comissão de juristas instituída para elaboração do anteprojeto, é preciso refletir sobre a possibilidade de expandir e incluir as alíneas *b* e *c*, podendo, assim, ser transformadas em um capítulo à parte do atual anteprojeto. A relatora entende que é difícil avançar em alguns termos no contexto nacional, entre eles o da defesa nacional e da segurança do Estado (2021, s.p.).

O texto do anteprojeto é um bom ponto de partida para o debate fundamental sobre a relevância de se regular uma participação mais ativa e transparente no controle e acesso às informações dos titulares aos seus dados, em especial no caso de dados sensíveis e biométricos, como o uso de tecnologias de reconhecimento facial. O texto estabelece requisitos e limitações aos usos admissíveis dos dados pessoais por parte das autoridades, cria obrigações de transparência a serem respeitadas pelos controladores de dados e prevê a elaboração de relatórios de impacto na ocasião do tratamento de dados pessoais sensíveis (ITS, 2020, p. 8). O documento prevê ainda a necessidade de os sistemas responsáveis por decisões automatizadas serem auditáveis, não discriminatórios e passíveis de comprovação acerca de sua precisão e grau de acurácia (ITS, 2020, p. 9).

Atualmente, o anteprojeto encontra-se na Câmara dos Deputados à espera de um parlamentar que o apresente formalmente, tornando-o, assim, um projeto de lei (PL). Após, o futuro PL seguirá os trâmites comuns do pro-

cesso legislativo, sendo submetido à avaliação das mais diversas comissões, votado, enviado ao Senado e submetido à sanção presidencial (Costa; Reis, 2021, s.p.).

Em razão de todo o exposto, percebe-se que o tratamento de dados pessoais no âmbito do controle migratório encontra-se em uma zona cinzenta. Dos três cenários apresentados no início do tópico, apenas o primeiro apresenta uma regulamentação específica, ou seja, o tratamento de dados pessoais realizado por companhias aéreas para a execução dos seus serviços deve seguir as orientações da Lei Geral de Proteção de Dados. Os outros dois cenários ainda não são regulamentados.

O segundo cenário, do tratamento de dados pessoais com o objetivo da segurança pública, deverá seguir as orientações do Anteprojeto da LGPD Penal, que ainda está em tramitação, e que, sem dúvidas, poderá sofrer modificações até a sua entrada em vigor. Por fim, o terceiro cenário, do tratamento voltado para a segurança nacional, ainda não possui uma legislação específica ou sequer um projeto de lei em andamento, configurando, portanto, uma lacuna no ordenamento jurídico brasileiro.

Por fim, compreende-se que a regulação estatal é condição urgente para sustentar o próprio papel evolutivo das tecnologias empregadas em sistemas de decisões automatizadas e garantir a defesa do humano ao revés dos danos que tais sistemas porventura possam causar.

CONSIDERAÇÕES FINAIS

O aumento exponencial dos fluxos migratórios e a utilização crescente do uso massivo das tecnologias digitais, para incrementar o controle nas fronteiras, trazem à tona o debate sobre a proteção dos dados pessoais dos migrantes. Nas iniciativas do Reino Unido e Canadá, pode-se notar a utilização e implementação de algoritmos em diferentes momentos do processo migratório. Uma das finalidades observadas seria aquela de acelerar e/ou facilitar os processos migratórios; entretanto, também é possível identificar a possibilidade de usos diversos desses sistemas automatizados, como a discriminação baseada em dados de origem, raça e gênero.

Trata-se de temáticas de extrema complexidade, que envolvem desde os movimentos migratórios, a coleta e tratamento de dados pessoais, as decisões automatizadas e a sociedade da vigilância, além dos avanços de novas tecnologias de perfilamento. Diante do exposto, compreende-se que as decisões automatizadas realizadas no âmbito de controle migratório não

apenas têm o potencial de discriminar migrantes, como provavelmente já o vem fazendo desde sua implementação. Possíveis soluções para que estes sistemas passem a atribuir um tratamento igualitário a migrantes, respeitando convenções internacionais que estabelecem a não discriminação e igualdade entre pessoas, consistiriam na maior transparência sobre as variáveis usadas para as decisões tomadas, bem como na inclusão e constante atualização de variáveis que não estivessem somente vinculadas a bases de dados de decisões pretéritas (Schippers; Gagliardi, 2021).

Neste contexto, compreende-se que o Direito precisa estar atento aos avanços tecnológicos, e, no que diz respeito à proteção de dados pessoais, é preciso considerar, além do princípio da dignidade da pessoa humana, a "finalidade, pertinência, proporcionalidade, simplificação, harmonização e necessidade" (Rodotà, 2008, p. 10).

Por fim, conclui-se a contribuição com uma perspectiva referente a novos desafios e novas abordagens: em um primeiro momento, compreende-se que é preciso buscar a criação de algoritmos isentos de vieses, que garantam a isonomia nas decisões automatizadas, inclusive de cunho migratório. Na sequência, entende-se que é necessário regular o tratamento dos dados pessoais no contexto migratório, em especial no que diz respeito ao tratamento de dados pessoais para segurança pública e para segurança nacional, considerando as lacunas da Lei Geral de Proteção de Dados e o andamento do Anteprojeto da LGPD Penal.

REFERÊNCIAS

ABDI. *Fronteira Tech*. 2021a. Disponível em: https://www.abdi.com.br/projetos/fronteira-tech. Acesso em: 20 set. 2021.

_____. *Fronteira Tech é lançado em Roraima*. 2021b. Disponível em: https://www.abdi.com.br/postagem/fronteiratech-e-lancado-em-roraima. Acesso em: 20 set. 2021.

ACNUR. *Dados sobre refúgio*. Publicado em: 18 jun. 2021. Disponível em: https://www.acnur.org/portugues/dados-sobre-refugio/. Acesso em: 17 set. 2021.

BAUMAN, Zygmunt; LYON, David. *Vigilância líquida*. Trad. Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013.

BANCO MUNDIAL. *World Development Report 2021*. Disponível em: https://wdr2021.worldbank.org/stories/crossing-borders/. Acesso em: 21 set. 2021.

BARNET. *Joint Council for the Welfare of Immigrants*. Disponível em: https://www.barnet.gov.uk/directories/directme/joint-council-welfare-immigrants. Acesso em: 30 nov. 2021.

BBC NEWS. *Home Office drops "racist" algorithm from visa decisions*. Publicado em: 4 ago. 2020. Disponível em: https://www.bbc.com/news/technology-53650758. Acesso em: 21 mar. 2021.

BRASIL. *Constituição da República Federativa do Brasil de 1988*. Disponível em: http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 7 dez. 2021.

_____. *Contrato RFB-Copol nº 22-2015*. Disponível em: http://receita.economia.gov.br/sobre/licitacoes-e-contratos/contratos-deti/2015/contrato-rfb-copol-no-22-2015-nec.pdf/view. Acesso em: 10 set. 2021.

_____. *Lei nº 12.527, de 18 de novembro de 2011*. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2011/Lei/L12527.htm. Acesso em:
20 set. 2021.

_____. *Lei nº 13.445, de 24 de maio de 2017*. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2017/lei/l13445.htm. Acesso em: 20 set. 2021.

_____. *Lei nº 13.709, de 14 de agosto de 2018*. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 23 set. 2021.

BRUNO, Fernanda. Entrevista: Fernanda Bruno. Vigilância hoje. *Revista Dispositiva*, v. 2, n. 1, maio 2013/out. 2013. Disponível em: http://periodicos.pucminas.br/index.php/dispositiva/article/download/6091/5680. Acesso em: 10 set. 2021.

COSTA, Eduarda; REIS, Carolina. *Histórico da LGPD Penal*: o que foi feito até aqui e quais são os próximos passos? Publicado em: 16 abr. 2021. Disponível em: https://lapin.org.br/2021/04/16/lgpd-penal-o-que-foi-feito-ate-aqui-e-quais-sao-os-proximos-passos/. Acesso em: 23 set. 2021.

COSTA, Frederico Carlos de Sá. Sobre o conceito de "segurança nacional". In: *Tensões Mundiais*, [s.l.], v. 5, n. 9, p. 123-140, 2018. DOI: 10.33956/tensoesmundiais.v5i9jul/dez.670. Disponível em: https://revistas.uece.br/index.php/tensoesmundiais/article/view/670. Acesso em: 7 dez. 2021.

DONEDA, Danilo. *Da privacidade à proteção de dados pessoais*. Rio de Janeiro: Renovar, 2006.

EUROPEAN COMMISSION. Article 29 Working Party. Guidelines on automated individual decision-making and profiling for the purposes of regulation 2016/679, 17/EN, WP251rev.01, Oct. 2017. Disponível em: https://ec.europa.eu/newsroom/article29/items/612053. Acesso em: 16 set. 2021.

FONTOURA, Natália de Oliveira; RIVERO, Patricia Silveira; RODRIGUES, Rute Imanishi. Segurança pública na Constituição Federal de 1988: continuidades e perspectivas. In: *IPEA*. Políticas sociais: acompanhamento e análise – Vinte anos da Constituição Federal, v. 3, 2009. Disponível em: https://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=5606. Acesso em: 7 dez. 2021.

FOXGLOVE. *Who we are*. Disponível em: https://www.foxglove.org.uk/who-we-are/. Acesso em: 30 nov. 2021.

FRAZÃO, Ana. *Discriminação algorítmica*: a relação entre homens e máquinas. Publicado em: 28 jul. 2021. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/discriminacao-algoritmica-a-relacao-entre-homens-e-maquinas-28072021. Acesso em: 16 set. 2021.

GEDIEL, José Antônio Peres; CORRÊA, Adriana Espíndola. Proteção de dados pessoais nos processos migratórios. In: BIONI, Bruno et al. *Tratado de proteção de dados pessoais*. Forense. Edição do Kindle. 2021.

GILLIOM, John; MONAHAN, Torin. *Supervision*: an introduction to the surveillance society. Chicago: The University of Chicago Press, 2013.

HEAVEN, Will Douglas. The UK is dropping an immigration algorithm that critics say is racist. *MIT Technology Review*. Publicado em: 5 ago. 2020. Disponível em: https://www.technologyreview.com/2020/08/05/1006034/the-uk-is-dropping-an-immigration-algorithm-that-critics-say-is-racist/. Acesso em: 21 mar. 2021.

HILDEBRANDT, Mireille. Defining profiling. A new type of knowledge. In: HILDEBRANDT, Mireille; GUTWIRTH, Serge. *Profiling the European Citizen*: cross-disciplinary perspectives, Londres: Springer, p. 17-44, 2008. Disponível em: https://www.researchgate.net/publication/226744267_Defining_Profiling_A_New_Typ e_of_Knowledge. Acesso em: 10 set. 2021.

IRIS. *Programa Embarque Seguro*: reconhecimento facial em aeroportos no Brasil. Publicado em: 2 dez. 2020. Disponível em: https://irisbh.com.br/programa-embarque-seguro-questionamentos-sobre-reconhecimento-facial-em-aeroportos-no-brasil/. Acesso em: 28 set. 2021.

ITS. *Comentários ao Anteprojeto de Lei de Proteção de Dados para a Segurança Pública*: tecnologia de reconhecimento facial. Disponível em: https://itsrio.org/pt/publicacoes/comentarios-ao-anteprojeto-de-lei-de-protecao-de-dados-para-a-seguranca-publica/. Acesso em: 20 set. 2021.

JUMPER, John; EVANS, Richard; PRITZEL, Alexander et al. Highly accurate protein structure prediction with AlphaFold. *Editorial Nature*, 596, 2021.

KAUFMAN, Dora. *A inteligência artificial irá suplantar a inteligência humana?* Barueri: Estão das Letras e Cores, 2019.

LAPIN. *Vigilância automatizada*: uso de reconhecimento facial pela Administração Pública. Publicado em: jul. 2021. Disponível em: https://lapin.org.br/2021/07/07/

vigilancia-automatizada-uso-de-reconhecimento-facial-pela-administracao-publica-no-brasil/. Acesso em: 20 set. 2021.

LINDEN, Greg; SMITH, Brent; YORK, Jeremy. Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing*, v. 7, 2003.

MAGRANI, Eduardo. *Entre dados e robôs*: ética e privacidade na era da hiperconectividade. 2. ed. Porto Alegre: Arquipélago Editorial, 2019.

MAHMOUDI, Matt. *Ban dangerous facial recognition technology that amplifies racist policing*. Publicado em: 26 jan. 2021. Disponível em: https://www.amnesty.org/en/latest/press-release/2021/01/ban-dangerous-facial-recognition-technology-that-amplifies-racist-policing/.

MARRAFON, Marco; MEDON, Filipe. *Importância da revisão humana das decisões automatizadas na Lei Geral de Proteção de Dados*. Publicado em: 9 set. 2019. Disponível em: https://www.conjur.com.br/2019-set-09/constituicao-poder-importancia-revisao-humana-decisoes-automatizadas-lgpd. Acesso em: 7 dez. 2021.

MCCARTHY, John et al. *A proposal for the Dartmouth summer conference on artificial intelligence* – Conference Announcement for the seminal meeting on AI, 1955. Disponível em: http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html. Acesso em: 24 out. 2021.

MCCULLOCH, Warren Sturgis; PITTS, Walter. *A logical calculus of the ideas immanent in nervous activity*. Grã-Bretanha: Bulletin of Mathematical Biophysics 5, 1943.

MENDES, Laura Schertel. *Anteprojeto de Lei de Proteção de Dados para Segurança Pública e Persecução Penal*: construção, temas e perspectivas. Publicado em: 26 abr. 2021. Disponível em: https://www.anymeeting.com/916-078-548/EA54DC89884C3F. Acesso em: 27 abr. 2021.

_____. *Decisão histórica do STF reconhece direito fundamental à proteção de dados pessoais*. 2020. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/decisao-historica-do-stf-reconhece-direito-fundamental-a-protecao-de-dados-pessoais-10052020. Acesso em: 20 set. 2021.

MINSKY, Marvin. *A Neural-Analogue Calculator Based upon a Probability Model of Reinforcement*. Massachusetts: Harvard University Psychological Laboratories, 1952.

MOLNAR, Petra; GILL, Lex. *Bots at the gate*: a human rights analysis of automated decision-making in Canada's immigration and refugee system. International Human Rights Program (Faculty of Law, University of Toronto) and the Citizen Lab (Munk School of Global Affairs and Public Policy, University of Toronto, 2018). Disponível em: https://citizenlab.ca/wp-content/uploads/2018/09/IHRP-Automated-Systems-Report-Web-V2.pdf. Acesso em: 21 mar. 2021.

MOZUR, Paul. Inside China's dystopian dreams: A.I., shame and lots of cameras. *The New York Times*. Publicado em: 7 ago. 2018. Disponível em: https://www.nytimes.com/2018/07/08/business/china-surveillance-technology.html. Acesso em: 22 set. 2021.

NAUMANN, Felix. *Data profiling revisited*. 2014. ACM SIGMOD Record, 42. Disponível em: https://dl.acm.org/doi/10.1145/2590989.2590995. Acesso em: 27 abr. 2021.

NEC. *Receita Federal utilizará tecnologia de identificação facial da NEC em 14 aeroportos internacionais do País*. 2016. Disponível em: https://br.nec.com/pt_BR/press/PR/20160409060302_11186.html. Acesso em: 10 set. 2021.

OLIVEIRA, Samuel R. de. *Sorria, você está sendo filmado!*: repensando direitos na era do reconhecimento facial. São Paulo: Thomson Reuters Brasil, 2021.

ONG, Thuy. *Dubai Airport is going to use face-scanning virtual aquariums as security checkpoints*. Publicado em: 10 out. 2017. Disponível em: https://www.theverge.com/2017/10/10/16451842/dubai-airport-face-recognition-virtual-aquarium. Acesso em: 22 set. 2021.

PEIXOTO, Fabiano Hartmann. Direito e inteligência artificial na (não) redução de desigualdades globais: decisões automatizadas na imigração e sistemas de refugiados. *Revista Direitos Culturais*, Santo Ângelo, v. 15, n. 37, set./dez. 2020. Disponível em: http://san.uri.br/revistas/index.php/direitosculturais/article/view/222/86. Acesso em: 26 mar. 2021.

QUOC, Viet Le et al. *Building high-level features using large scale unsupervised learning*. In: Proceedings of the 29th International Conference on International Conference on Machine Learning (ICML), 2012. Edimburgo, Escócia.

RAVURI, Suman; LENC, Karel; WILLSON, Matthew et al. Skilful precipitation nowcasting using deep generative models of radar. *Editorial Nature*, 597, p. 672, 2021.

RICH, Elaine; KNIGHT, Kevin. *Artificial intelligence*. 2. ed. Nova Iorque: McGraw-Hill, 1991.

RODOTÀ, Stefano. *A vida na sociedade da vigilância*: a privacidade hoje. Organização, seleção e apresentação de Maria Celina Bodin de Moraes. Trad. Danilo Doneda e Luciana Cabral Doneda. Rio de Janeiro: Renovar, 2008.

_____. Transformações do corpo. *Revista Trimestral de Direito Civil*, v. 19, n. 5, p. 91-107, 2004.

ROSENBLATT, Frank. *Principles of neurodynamics*: perceptions and the theory of brain mechanism. Buffalo: Cornell Aeronautical Lab Inc., 1961.

SCHIPPERS, Laurianne-Marie; GAGLIARDI, Marília Papaléo. *Inteligência artificial e controle migratório*: algoritmos podem discriminar migrantes? Publicado em: 16 jan. 2021. Disponível em: https://medium.com/o-centro-de-ensino-e-pesquisa-em-

inova%C3%A7%C3%A3o-est%C3%A1/intelig%C3%AAncia-artificial-e-controle-migrat%C3%B3rio-algoritmos-podem-discriminar-migrantes-85d04d152440. Acesso em: 10 set. 2021.

SEI. *Barramento de Serviços do PEN*. Disponível em: https://portalsei.df.gov.br/barramento/. Acesso em: 7 dez. 2021.

SENADO. *PEC que inclui a proteção de dados pessoais na Constituição volta para o Senado*. Publicado em: 3 set. 2021. Disponível em: https://www12.senado.leg.br/noticias/materias/2021/09/03/pec-que-inclui-a-protecao-de-dados-pessoais-na-constituicao-volta-para-o-senado. Acesso em: 20 set. 2021.

SERPRO. *Embarque + seguro*. Disponível em: https://campanhas.serpro.gov.br/embarque-mais-seguro/. Acesso em: 20 set. 2021.

SOLOVE, Daniel J. *Understanding privacy*. Cambridge: Harvard University Press, 2008.

SOUZA, Bruno. *Panóptico: reconhecimento facial renova velhas táticas racistas de encarceramento.* Publicado em: 22 abr. 2021. Disponível em: http://observatorioseguranca.com.br/panoptico-reconhecimento-facial-renova-velhas-taticas-racistas-de-encarceramento/. Acesso em: 6 dez. 2021.

UNIÃO EUROPEIA. Parlamento e Conselho. Regulamento (EU) nº 2016/679, de 27 de abril de 2016. Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva nº 95/46/CE (Regulamento Geral de sobre Proteção de Dados). *Jornal Oficial da União Europeia,* [s.l.], L 119/1, 4 maio 2016. Disponível em: https://eurlex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434. Acesso em: 10 set. 2021.

UN, Desa. *International Migration 2020 Highlights*. Publicado em: jan. 2021. Disponível em: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/international_migration_2020_highlights_ten_key_messages.pdf. Acesso em: 17 set. 2021.

UNODC. *How Covid-19 restrictions and the economic consequences are likely to impact migrant smuggling and cross-border trafficking in persons to Europe and North America*. Disponível em: https://www.unodc.org/documents/islamicrepublicofiran//2020/05/Covid-related-impact-on-SoM-TiP.PDF. Acesso em: 22 set. 2021.

VIOLA, Mario; HERINGER, Leonardo; CARVALHO, Celina. *O anteprojeto da LGPD Penal e as regras sobre transferência internacional de dados pessoais*. Publicado em: ago. 2021. Disponível em: https://itsrio.org/wp-content/uploads/2021/07/Relatorio-Transferencia-de-dados-pessoais.pdf. Acesso em: 20 set. 2021.

ZUBOFF, Shoshona. *A era do capitalismo de vigilância*: a luta por um futuro humano na nova fronteira do poder. Trad. George Schlesinger. 1. ed. Rio de Janeiro: Intrínseca, 2020.

Sobre as autoras:

Stéfani Reimann Patz | *E-mail:* stefani.patz@hotmail.com

Mestranda em Direitos Especiais pelo Programa de Pós-Graduação *Stricto Sensu* em Direito. Mestrado e Doutorado da Universidade Regional Integrada do Alto Uruguai e das Missões (URI), *Campus* Santo Ângelo/RS. Bolsista Capes/Prosuc. Bacharela em Direito pela URI, *Campus* Santo Ângelo/RS. Pesquisadora voluntária dos projetos de pesquisa "Crisálida: Direito e Arte", "Internet, liberdade de informação, manipulação de comportamentos e a desestabilização do processo democrático" e do Centro de Estudos e Pesquisas em Direito e Tecnologia (Cedetec). Membro do Instituto Nacional de Proteção de Dados (INPD).

Thami Covatti Piaia | *E-mail:* thamicovatti@hotmail.com

Doutora em Direito pela Universidade Federal do Rio Grande do Sul – UFRGS (2013). Contemplada com bolsa da Capes durante o período de Doutoramento e contemplada com bolsa do Programa de Doutorado Sanduíche no Exterior (PDSE) pelo período de onze meses na Universidade de Illinois, *Campus* de Urbana-Champaign – Estados Unidos. Mestre em Direito pela Universidade Regional Integrada do Alto Uruguai e das Missões – URI, *Campus* de Santo Ângelo/RS. Graduada em Direito pela Universidade Regional Integrada do Alto Uruguai e das Missões – URI, *Campus* de Frederico Westphalen/RS. Professora na Graduação e no Programa de Pós-Graduação *Stricto Sensu* em Direito. Mestrado e Doutorado da Universidade Regional Integrada do Alto Uruguai e das Missões – URI, *Campus* de Santo Ângelo/RS. Possui inscrição na Ordem dos Advogados do Brasil.

Data de submissão: 30 de setembro de 2021.
Data de aceite: 7 de dezembro de 2021.