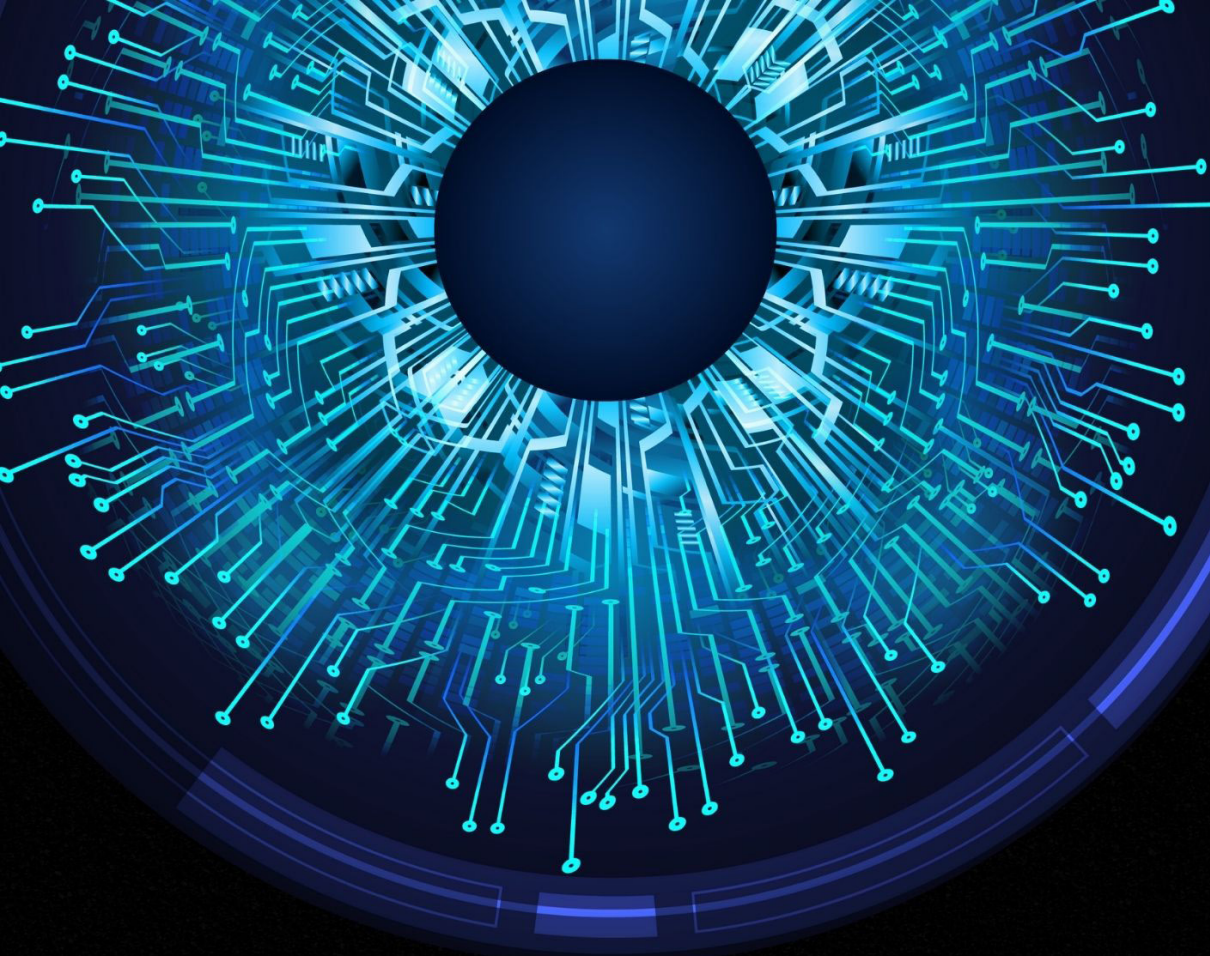




INTELIGÊNCIA ARTIFICIAL NO SUL GLOBAL:

**REGULAÇÃO, RISCOS DISCRIMINATÓRIOS,
GOVERNANÇA E RESPONSABILIDADES**

**Haide Maria Hupffer
Wilson Engelmann
Taís Fernanda Blauth
(organizadores)**

Este livro é o resultado da pesquisa e das relações interinstitucionais produzidas no âmbito do seguinte projeto de investigação científica:

INTELIGÊNCIA ARTIFICIAL E SOCIEDADE DE ALGORITMOS: REGULAÇÃO, RISCOS DISCRIMINATÓRIOS, GOVERNANÇA E RESPONSABILIDADES

Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul - FAPERGS

Proc. n.º 61271.674.22274.26112021

Edital Fapergs 07/2021

Programa Pesquisador Gaúcho – PqG



**HAIDE MARIA HUPFFER
WILSON ENGELMANN
TAÍS FERNANDA BLAUTH
(ORGANIZADORES)**

**INTELIGÊNCIA
ARTIFICIAL
NO SUL GLOBAL:
REGULAÇÃO, RISCOS DISCRIMINATÓRIOS,
GOVERNANÇA E RESPONSABILIDADES**

CASA LEIRIA
SÃO LEOPOLDO/RS
2023

**INTELIGÊNCIA ARTIFICIAL NO SUL GLOBAL: REGULAÇÃO,
RISCOS DISCRIMINATÓRIOS, GOVERNANÇA E
RESPONSABILIDADES**

DOI: 10.29327/5312711

Organizadores: Haide Maria Hupffer,
Wilson Engelmann e
Taís Fernanda Blauth.

Revisão e edição: Casa Leiria.

Os textos são de responsabilidade de seus autores.

Qualquer parte desta publicação pode ser reproduzida,
desde que citada a fonte.

Casa Leiria Conselho Editorial

Ana Carolina Einsfeld Mattos	(UFRGS)
Ana Patrícia Sá Martins	(UEMA)
Antônia Sueli da Silva Gomes Temóteo	(UERN)
Glícia Marili Azevedo de Medeiros Tinoco	(UFRN)
Haide Maria Hupffer	(Feevale)
Isabel Cristina Arendt	(Unisinos)
Isabel Cristina Michelin de Azevedo	(UFS)
José Ivo Follmann	(Unisinos)
Luciana Paulo Gomes	(Unisinos)
Luiz Felipe Barboza Lacerda	(UNICAP)
Márcia Cristina Furtado Ecoten	(Unisinos)
Rosangela Fritsch	(Unisinos)
Tiago Luís Gil	(UnB)

161 Inteligência artificial no sul global: regulação, riscos discriminatórios,
governança e responsabilidades [recurso eletrônico] / organização
Haide Maria Hupffer, Wilson Engelmann, Taís Fernanda Blauth.
São Leopoldo: Casa Leiria, 2023.

Disponível em: <[http://www.guaritadigital.com.br/casaleiria/acervo/
direito/ianosulglobal/index.html](http://www.guaritadigital.com.br/casaleiria/acervo/direito/ianosulglobal/index.html) >

ISBN 978-85-9509-096-5

1. Direito – Inteligência artificial – Regulação. 2. Inteligência artificial –
Direito – Governança. III. Inteligência artificial – Direito – Ética. I. Hupffer,
Haide Maria (Org.). II. Engelmann, Wilson (Org.). III. Blauth, Taís Fernanda
(Org.).

CDU 34:004.8

**INTELIGÊNCIA
ARTIFICIAL
NO SUL GLOBAL:
REGULAÇÃO, RISCOS DISCRIMINATÓRIOS,
GOVERNANÇA E RESPONSABILIDADES**

SUMÁRIO

- 9 **Prefácio**
Denis Coitinho
- 13 **Inteligência Artificial responsável: significados e desafios**
Wilson Engelmann
- 35 **Inteligencia Artificial y regulación dentro del Marco Constitucional Argentino: algunas reflexiones sobre el impacto de las nuevas tecnologías**
Martín Julián Acevedo – Miño
- 45 **Discriminação algorítmica e colonialismo digital**
José Luiz de Moura Faleiros Júnior
- 75 **Inteligência Artificial e discriminação algorítmica: marcos regulatórios e parâmetros éticos**
Haide Maria Hupffer
Gustavo da Silva Santanna
- 111 **A nova dimensão dos neurodireitos cibernéticos: imunização pactuada contra a ciberescravidão?**
Paola Cantarini-Guerra
Willis Santiago Guerra Filho
- 137 **A Inteligência Artificial dos *Large Language Models (LLMs)* e os riscos aos Direitos Autorais: diretrizes aplicadas às plataformas e novos deveres éticos-jurídicos para sua utilização**
Cristiano Colombo
Guilherme Damasio Goulart
- 171 **Violencias en las redes sociales: la competencia desigual por su sentido**
Nelson Arteaga Botello

- 193 Inteligência Artificial Generativa frente ao dever de fundamentar as decisões judiciais: o ChatGPT e os limites do artigo 489, §1º, do CPC
Darci Guimarães Ribeiro
Afonso Vinício Kirschner Fröhlich
- 221 Inteligencia Artificial y teoría de la evolución. Reflexiones en torno a la viabilidad de la *global lex digitalis*
Carlos de Cores Helguera
- 239 Consecuencias jurídicas del uso de ChatGPT en la acción de tutela en Colombia
Sebastián Díaz Bolívar
Dany Steven Gómez Agudelo
- 259 Segurança Pública em Cidades Inteligentes sob uma perspectiva local
José Jorge Barreto Torres
Patrícia Verônica Nunes Carvalho Sobral de Souza
Diogo de Calasans Melo Andrade
- 279 Meio Ambiente Digital, Inteligência Artificial e Educação
Alexandre Cortez Fernandes
Patrícia Montemeyzo
- 301 Sobre os autores
- 307 Índice remissivo

PREFÁCIO

A tecnologia orientada por Inteligência Artificial (IA), entendida como a capacidade de um sistema – como um software ou incorporada em um aparelho – para executar tarefas comumente associadas a seres inteligentes, já é uma constante em nossas vidas, seja pelas recorrentes recomendações de filmes que recebemos pelos serviços de *streaming*, tal como Netflix e Amazon Prime, seja pelos filtros de e-mails que usamos, ou pela utilização de algum aplicativo de navegação por GPS, como o Waze, e ela não parece problemática em muitas áreas. Ao contrário, ela parece facilitar nossa vida. Entretanto, esta tecnologia está se estendendo progressivamente para certos domínios nos quais, provavelmente, terá um impacto maior, assim como decidir em circunstâncias de risco, estabelecer prioridades entre pessoas e fazer julgamentos complexos e, portanto, terá que tomar decisões que já se enquadram no domínio moral. Por isso, parece importante pensar sobre os algoritmos que alimentam esses produtos. Os carros autônomos, por exemplo, precisarão tomar decisões sobre como distribuir o risco entre os passageiros, pedestres e ciclistas, isto é, entre os que utilizam as vias públicas. As armas autônomas letais, terão que identificar e selecionar os alvos humanos que serão eliminados. Por sua vez, algoritmos que já estão em uso no sistema de saúde e no sistema judiciário em alguns países, estabelecem a prioridade de quem receberá um transplante de órgão, bem como aconselham os juízes sobre quem deve obter liberdade condicional ou uma sentença maior de prisão. Até mesmo um aplicativo inocente como Spotify pode apresentar um problema ético em seu algoritmo, como sexismo. Uma pesquisa conduzida por A. Ferraro, X. Serra e C. Bauer demonstrou que o algoritmo da plataforma privilegia músicos homens, sendo as artistas mulheres pouco sugeridas. Quando testaram o algoritmo descobriram que, em média, as seis primeiras faixas recomendadas são de homens (Break the Loop: Gender Imbalance in Music Recommenders, *CHIIR '21*, 2021).

A promessa da IA é que ela melhore nossa vida, pois esses algoritmos podem, teoricamente, decidir sem vieses de gênero, raça ou classe, de forma lógica e racional, evitando qualquer parcialidade e discriminação. Entretanto, pode ser que isso não ocorra tal como esperado, uma vez que todas essas decisões referidas anteriormente, inevitavelmente, parecem envolver padrões éticos e avaliações morais complexas, o que pode resultar em injustiças. Por exemplo, os carros autônomos devem sempre se esforçar para minimizar as baixas, mesmo que às vezes isso signifique sacrificar seus próprios passageiros para um bem maior? As armas autônomas letais devem sempre objetivar a vitória, mesmo ao custo de eliminar um alvo civil ou um soldado ferido, o que colocaria em risco a dignidade humana? As crianças devem sempre ter prioridade para transplantes de órgãos, mesmo quando um paciente mais velho é uma combinação genética melhor para um órgão disponível? Os algoritmos usados em tribunais devem sempre procurar reduzir a reincidência, mesmo resultando em uma discriminação injusta para os réus negros?

O problema que nos deparemos atualmente é que alguns usos da IA utilizados em diferentes contextos, como no trânsito, educação, saúde, judiciário, política etc., a exemplo dos programas de reconhecimento facial, avaliação de currículos, classificação de fotografias por assunto e programas que influenciam as decisões judiciais no âmbito penal, tal como o COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), têm revelado disparidades preocupantes de desempenho com base na raça e no gênero das pessoas. Essas disparidades levantam questões urgentes sobre como o uso da IA pode funcionar para promover a justiça ou consolidar a injustiça, podendo revelar, por exemplo, discriminação institucional, injustiça estrutural, como racismo, sexismo e preconceito de classe. Em uma era de crescente utilização da IA, a ideia de que o progresso tecnológico conduz necessariamente à emancipação oportuniza uma grande expectativa. Mas isto pode ser ilusório, uma vez que os sistemas de IA são ainda ferramentas com as quais as pessoas podem exercer poder umas

sobre as outras. E, como usual, podemos exercer o poder para promover a emancipação ou a subjugação humana.

Importante frisar que a importância deste debate se percebe também nas novas agendas políticas e sociais que, inclusive, tem se estabelecido em documentos de caráter internacional, como a “Agenda 2030” da ONU, e outros tantos documentos internacionais, em que os Direitos Humanos e a justiça social aparecem como eixos fundamentais para a garantia do futuro da sociedade humana, inclusive ponderando sobre os impactos da tecnologia em nossa vida comum, especialmente nos impactos que recaem nos cidadãos socialmente mais vulneráveis. E essa preocupação parece muito adequada, uma vez que a aplicação da IA parece impactar mais a vida dos cidadãos mais vulneráveis, e isso constitui um claro desrespeito aos direitos humanos, critério este que deve orientar normativamente as sociedades que se pretendem democráticas.

Com essa urgência em mente, queremos saudar a publicação do livro *Inteligência Artificial no Sul Global: Regulação, Riscos Discriminatórios, Governança e Responsabilidades*, organizado pelos professores e pesquisadores Haide Maria Hupffer (Feevale), Wilson Engelmann (Unisinus) e Taís Fernanda Blauth (University of Groningen), que trata com maestria deste relevante problema contemporâneo. O livro investiga importantes temas relativos à inteligência artificial, tais como a questão de regulação, o problema da responsabilidade moral e jurídica, a discriminação algorítmica e colonialismo digital, os neurodireitos cibernéticos, a proteção aos direitos humanos, os riscos aos direitos autorais e à privacidade, o fundamento das decisões judiciais, a educação no contexto da IA, entre outros. A obra é uma coletânea de textos que contou com o financiamento público da FAPERGS (Fundação de Amparo à Pesquisa do Estado do RS) e que reúne uma gama diversificada e qualificada de pesquisadores da área do direito, pesquisadores tanto nacionais quanto internacionais, reunidos pela ligação com a temática. Um ponto central que parece unificar os vários capítulos do livro é o alerta de que devemos enfrentar o desafio enquanto sociedade de regulamentar a IA para

garantir a promoção de igualdade dos cidadãos, impedindo qualquer discriminação.

Dois pontos merecem destaque. Primeiro, a sugestão de orientarmos a IA seguindo certos princípios éticos, identificados a partir dos princípios propostos pelo *Chinese National New Generation Artificial Intelligence Governance Professional Committee*, tais como Harmonia e Amizade, Equidade e Justiça, Tolerância, Respeito à Privacidade, Segurança, Responsabilidades compartilhadas, Colaboração e Governança, bem como levando em conta as exigências formuladas pelo *European High-Level Expert Group in AI*, como o Bem-estar social e ambiental, a Diversidade, Não-discriminação e Equidade, a Supervisão humana, a Privacidade e Governança de dados, a Robustez técnica e Segurança do sistemas de IA, a Transparência e a Responsabilidade. Segundo, alertar para o perigo da discriminação algorítmica, objetivando identificar algumas externalidades negativas e os reflexos que algoritmos enviesados podem produzir na sociedade e aos direitos fundamentais, bem como examinar os principais movimentos regulatórios para lidar com as difíceis questões éticas, sociais e políticas sobre os riscos da discriminação algorítmica para que a IA não perpetue o racismo estrutural enraizado nas decisões humanas.

O aspecto valioso neste trabalho é que ele faz uma defesa intransigente a respeito de como devemos estruturar modelos de regulação para assegurar a IA responsável, propondo que sigamos certos princípios jurídicos e morais como estruturas normativas para orientação destas novas tecnologias, tais como justiça e equidade, o respeito à privacidade, a garantia da segurança, a robustez e também, fundamentalmente, a transparência. Com isso, espera-se impedir que o desenvolvimento tecnológico, que é desejável para todos nós, tenha um caráter predominantemente econômico, tendo por foco a centralidade humana.

Denis Coitinho

Coordenador Executivo e Professor do
Programa de Pós-Graduação em Filosofia da Unisinos

INTELIGÊNCIA ARTIFICIAL RESPONSÁVEL: SIGNIFICADOS E DESAFIOS¹

Wilson Engelmann²
DOI: 10.29327/5312711.1-1

Resumo: Os avanços observados com a Inteligência Artificial (IA) são constantes e crescentes. Se objetiva estudar os desafios que a IA está trazendo, notadamente para o campo regulatório. O problema a ser respondido é: como estruturar modelos de regulação para assegurar a Inteligência Artificial responsável? A pesquisa foi realizada a partir da pesquisa bibliográfica, no Portal de Periódicos da Capes, e documental, com preferência a publicações realizadas a partir de 2018, usando-se as palavras-chave abaixo indicadas. Como resultados se destaca a importância dos princípios jurídicos para a estruturação de modelos regulatórios que possam ter aplicação nacional e global, destacadamente a partir da comparação entre China e União Europeia. Também se verifica a importância da teoria

- 1 Este trabalho é o resultado parcial das pesquisas realizadas pelo autor no âmbito dos seguintes projetos de pesquisa: a) Chamada CNPq n. 09/2020 – Bolsas de Produtividade em Pesquisa – PQ, projeto intitulado: “Percurso para ressignificar a Teoria Geral das Fontes do Direito: o Sandbox regulatório como um elemento estruturante da comunicação reticular entre o Direito e as nanotecnologias”; b) “Sistema do Direito, novas tecnologias, globalização e o constitucionalismo contemporâneo: desafios e perspectivas”, Edital Fapergs/Capes 06/2018 – Programa de Internacionalização da Pós-Graduação no RS. Este trabalho também está vinculado às pesquisas realizadas pelo autor nos seguintes Centros Internacionais de Investigação: CEDIS – Centro de Investigação & Desenvolvimento sobre Direito e Sociedade, da Faculdade de Direito da Universidade Nova de Lisboa, Portugal; do Instituto Jurídico Portucalense, da Universidade Portucalense, cidade do Porto, Portugal; e do CEAD – Centro Francisco Suárez – Centro de Estudos Avançados em Direito da Universidade Lusófona de Lisboa, Portugal.
- 2 Doutor e Mestre em Direito Público, Programa de Pós-Graduação em Direito da Universidade do Vale do Rio dos Sinos – Unisinos, Brasil; realizou Estágio de Pós-Doutorado em Direito Público-Direitos Humanos, no Centro de Estudos de Seguridad (CESEG) da Universidade de Santiago de Compostela, Espanha; Professor e Pesquisador do Programa de Pós-Graduação em Direito – Mestrado e Doutorado e do Mestrado Profissional em Direito da Empresa e dos Negócios, ambos da Unisinos; Bolsista de Produtividade em Pesquisa do CNPq; Líder do Grupo de Pesquisa JusNano. ORCID: <https://orcid.org/0000-0002-0012-3559>. Lattes: <http://lattes.cnpq.br/7143561813892945> E-mail: wengelmann@unisinos.br.

sobre os Direitos Humanos, a partir de John Finnis, como referencial estruturante à arquitetura autorregulatória regulada e principiológica para os avanços e desafios trazidos pela IA, a fim de se promover as bases da IA responsável.

Palavras-chave: Inteligência Artificial. Autorregulação Regulada. Princípios. Direitos Naturais. Direitos Humanos. Direitos Fundamentais. Comparação principiológica entre China e União Europeia.

Sumário: 1. Introdução. 2. Conhecendo a Inteligência Artificial. 3. Um plano de vida coerente e a *phrónesis* aristotélica: estruturas metodológico-éticas para regular a IA. 4. Os princípios jurídicos como estruturas normativas para desenvolver a chamada IA responsável. 5. Considerações Finais. Referências.

1. Introdução

O desenvolvimento da Inteligência Artificial é um dado da realidade atual, com projeções bastante expressivas de crescimento e utilização nas variadas áreas de atuação do ser humano. Essa característica levanta uma preocupação: o ser humano será substituído em suas atividades pelos sistemas de Inteligência Artificial? Como se deverá regular esses avanços, assegurando-se o desenvolvimento da chamada Inteligência Artificial responsável? Esse o problema que o artigo pretende responder. O tema é atual e relevante dado o crescimento das variadas utilizações da IA, por um lado; e a ausência de uma regulação específica, por outro lado. Segundo estudo, elaborado pelo McKinsey Technology Trends Outlook (2022), publicado em agosto de 2022, existem quatorze tendências em IA que estarão em alta a partir de 2023, se destacando, por exemplo, a IA aplicada; conectividade avançada; futuro da bioengenharia; futuro da mobilidade e o futuro da energia limpa, dentre outras aplicações. Esse cenário promissor e desafiador, concomitantemente, sublinha a importância do tema e da discussão sobre a chamada IA responsável.

O objetivo geral do trabalho é: estudar os mecanismos para regular os avanços da IA responsável. Os objetivos específicos são: a) conhecer as características e os elementos estruturantes da IA; b) estudar, a partir de John Finnis, o denominado *plano de vida coerente* e a *phrónesis* aristotélica, como estruturas metodológico-éticas para regular a IA; c) conhecer os princípios jurídicos, publicados em documentos normativos pela China e União Europeia, como estruturas normativas para desenvolver a chamada IA responsável.

Metodologicamente o trabalho foi desenvolvido a partir da pesquisa bibliográfica, principalmente por meio de pesquisa realizada no Portal de Periódicos da Capes, acessado a partir do site principal da Biblioteca da Unisinos, e pesquisa documental. Todas as fontes foram selecionadas a partir da utilização das palavras-chave que apresentam este trabalho, priorizando-se publicações realizadas a partir de 2018.

Para se desenvolver os objetivos – geral e específicos – acima apresentados, o trabalho está dividido em três capítulos, conforme se poderá ler a seguir.

2. Conhecendo a Inteligência Artificial

A criatividade do ser humano poderá levá-lo à fronteira do conhecimento, gerando novidades e, simultaneamente, riscos imprevisíveis, talvez, no limite, incontroláveis. Essas perspectivas não significam que se deverá abondar o seu espírito inquieto e criativo. Pelo contrário, será preciso avançar com cuidado e respeitando certos balizadores que são apresentados pela própria natureza ao ser humano, como manter a saúde do corpo e mente, e preservar o meio ambiente em que se vive.

Uma dessas áreas se chama de “Inteligência Artificial” (IA, doravante), que se posiciona em um espaço de convergência da Computação Aplicada e outros quadrantes tecnológicos com domínios científicos do Direito e da Filosofia, principalmente, mas não exclusivamente. Se observa gradativamente que as áreas “duras” ou “exatas”, apesar de serem aquelas onde a IA se desenvolve e se torna possível, precisa da “ajuda” dos conhecimentos gerados nos domínios menos técnicos e

mais humanos ou vinculados às Ciências Sociais Aplicadas e Ciências Humanas.

Um exemplo desse cruzamento de áreas poderá ser colhido na época histórica próxima da pandemia da COVID 19. Vale dizer, um artigo publicado na *Revista Nature*, de 23 de março de 2021, dizia em seu título que a “Ciência não seria suficiente para salvar a humanidade”. Trazendo o conteúdo do artigo para a abordagem sobre a IA, se deve problematizar uma questão-chave: quem está sendo chamado para desenvolver e avaliar os riscos que poderão ser gerados pelo uso em larga escala da IA? Via de regra, se buscam conselhos e/ou pareceres de especialistas, ou seja, de pessoas em Ciência, Tecnologia, Engenharia e Matemática (a sigla dessas áreas em inglês é STEM) – apesar de estar claro, desde o início da projeção sobre IA, que o comportamento humano, as motivações e a cultura serão fundamentais para uma resposta eficaz. Essa abordagem precisa mudar. A Ciência viabiliza os sistemas de IA, mas as atividades ou áreas inseridas nas Ciências Sociais Aplicadas, Humanidades e Artes para as pessoas e a Economia (a sigla dessas áreas em inglês é SHAPE) ajudam a chegar às realidades sociais, além da análise e avaliação dos impactos sociais, éticos, jurídicos e econômicos dos avanços a partir da IA. A percepção da humanidade é mais robusta quando STEM e SHAPE se unem. Por isso, se poderá dizer que a IA, como um gênero, representa fenômenos sociais, biológicos e tecnológicos, concomitantemente. A partir desse cenário, não se deverá focar excessivamente em evidências de ensaios de controle randomizados, com as variáveis do funcionamento da IA controladas em laboratórios, mas levar em conta as evidências observacionais e qualitativas nas quais as Ciências SHAPE estão imersas (adaptado a partir de Shah, 2021).

Qual o fundamento desse cruzamento de conhecimentos das áreas STEM e SHAPE? Pelas diversas facetas que se encontra no próprio conceito de IA, que pode ser entendida como a

simulación de procesos de inteligencia humana por medio de máquinas que se extiende a través de capacidades tales como el reconocimiento de voz, la toma de decisiones, la búsqueda semántica y las diversas técnicas de aprendizaje automático (Devang et al, 2019).

Por meio do sistema de IA se poderá simular características próprias do ser humano, como raciocinar, conectar dados e projetar resultados e tomar decisões. Por isso, a IA também se define como a “[...] *capacidad de un sistema para interpretar correctamente los datos externos y aprender de ellos para utilizarlos con el fin de alcanzar objetivos específicos*” (Cuervo Sánchez, 2021). Todas essas características são muito próximas daquelas que os seres humanos também apresentam e que, até agora, distinguia-os dos demais seres vivos que habitam o Planeta Terra. E das máquinas também. Será que essa distinção estaria com os seus dias contados? Não se sabe, ainda.

E tem mais um detalhe: para que o sistema de IA possa desenvolver todas essas “competências” precisa do chamado “algoritmo”, que é uma coisa pequena e simples; uma regra usada para automatizar o tratamento de um dado (Barbosa, 2021). Uma espécie de receita: “se isso acontecer, então faça *b*; se não, então faça *c*”. Tal a lógica “se/então/mais” da computação clássica. Entretanto, essa é a concepção original, pois

[...] surgiu um significado mais ambíguo, indicando qualquer sistema de *software* de tomada de decisão complexo e grande; qualquer meio de obter um conjunto de dados de entrada – e avaliá-los rapidamente, de acordo com um determinado conjunto de critérios (ou “regras”) (Smith, 2018).

O sistema complexo que representa o algoritmo viabiliza a Inteligência Artificial (IA) por meio do armazenamento de dados e, a partir deles, consegue estruturar séries de dados, que no Direito podem gerar a “jurimetria”, por exemplo, como uma ferramenta para se ter uma ideia do modo como decide determinado tribunal, a partir de dados coletados por um longo período, que são tratados, estruturados e sistematizados, permitindo-se detectar tendências no sentido de julgamentos e temas escolhidos (Smith, 2018). A partir daí se pode referir que existe certa superioridade da máquina – do sistema – em relação ao ser humano? Ainda não se tem uma resposta acabada para essa pergunta, mas se observa uma tendência para uma resposta positiva.

Algoritmos não são novos. Por décadas, eles têm servido como componentes integrais de todos os programas de

computador. Na atualidade, os algoritmos, por meio dos avanços do aprendizado de máquina, estão criando uma sociedade amplamente automatizada, transformando muitas facetas da vida. Muitos produtos e serviços – incluindo filtros de e-mail como o “spam”, diagnósticos médicos, marketing de produto, e carros autônomos – dependem de algoritmos de aprendizado de máquina e sua capacidade de fornecer poder e velocidade surpreendentes. Os algoritmos atuais são “robôs” digitais que possuem habilidades efetivamente autônomas de adaptação e aprendizado (adaptado a partir de Smith, 2018).

3. *Um plano de vida coerente e a phrónesis aristotélica: estruturas metodológico-éticas para regular a IA*

Esse panorama abre diversas frentes para a pesquisa, especialmente para se chegar a um ponto de equilíbrio na denominada “IA responsável”. Como se poderá construir essa possibilidade? A seguir se levantam algumas questões que deverão ser levadas em consideração. Os computadores podem ser treinados para detectar padrões nos dados, mesmo padrões tão sutis que os humanos podem não perceber. Entretanto, quando se trata de identificar a relação entre causa e efeito, as máquinas geralmente ficam perdidas, pois elas carecem de uma compreensão de bom senso sobre o funcionamento do mundo, observável pelo ser humano na sua vida cotidiana. Para que os computadores realizem qualquer tipo de tomada de decisão, eles precisarão entender a causalidade. A incorporação de modelos de causa e efeito em algoritmos de aprendizado de máquina também pode ajudar máquinas autônomas móveis a tomar decisões sobre como navegar no mundo. Mesmo algo como o ChatGPT, o popular gerador de linguagem natural que produz texto que parece ter sido escrito por um humano, poderia se beneficiar da incorporação da causalidade. Com ela, o ChatGPT se poderia construir um plano coerente para o que estava tentando dizer e garantir que fosse consistente com os fatos, tais quais são conhecidos (adaptado a partir de Savage, 2023). A mescla da IA nesse novo chat também provocam desafios para as atividades jurídicas, pois trazem as chamadas “ferramentas de IA

generativas”. Essa nova categoria da IA está tendo um momento cultural e comercial, sendo apontada como o futuro da pesquisa, gerando disputas legais sobre direitos autorais e causando pânico em escolas e universidades. A tecnologia, que usa grandes conjuntos de dados para aprender a gerar imagens ou textos que pareçam naturais, pode ser uma boa opção para o setor jurídico, que depende fortemente de documentos e precedentes padronizados, como o caso dos contratos, além da geração automatizada de documentos jurídicos, que tem sido uma área de crescimento há décadas, mesmo em dias de tecnologia baseada em regras, pois os escritórios de advocacia podem recorrer a grande quantidade de modelos altamente padronizados e bancos de precedentes para estruturar a geração de documentos, tornando os resultados muito mais previsíveis do que com a maioria saídas de texto livre (Stokel-Walker, 2023).

Essas possibilidades parecem surpreendentes, pois a atividade da advocacia, como a elaboração de peças processuais e outros documentos, e a produção de textos, até agora, eram atividades intelectuais produzidas por seres humanos. A mobilização da linguagem e da criatividade, produziam contratos, peças processuais, textos para jornais, artigos científicos, dentre tantas outras produções intelecto-textuais. Se aprendia desde cedo na Escola a construção de diversas estágios para se aprimorar a escrita, a leitura e a compreensão textual (Floridi, 2023). Os avanços multifacetados das aplicações da IA geram curiosidade sobre quais novidades ainda virão. Paralelamente, se deverá investigar como regular tantas novidades e gerir os riscos que poderão decorrer desses avanços. Se observa uma euforia com as interações viabilizadas por meio da IA. Esse fenômeno poderá provocar um velamento dos impactos éticos, jurídicos e sociais.

Se observa que a IA sempre se vincula a um *processo de aprendizagem*, estruturada a partir de dados: quanto maior o número de dados que o sistema tem à sua disposição, maior o seu nível de aprendizado. Nesse cenário, cabe destacar a Proposta de Regulamento do Parlamento Europeu e do Conselho Europeu (chamada de *AI Act*) que apresenta a categoria conceitual de *sistema de inteligência artificial*, como um *software*, desen-

volvido a partir de um variado *conjunto de técnicas e abordagens* orientadas para um determinado conjunto de *objetivos definidos pelo ser humano*, gerando saídas como conteúdo, previsões, recomendações ou decisões que influenciam os ambientes com os quais eles interagem (Regulamento, 2021). Essa concepção evidencia a importância da participação do ser humano na definição dos objetivos que devem ser alcançados pelo sistema de IA. Se tem uma característica fundamental: *a tecnologia deverá ficar sob o controle humano* (Codagnone et al, 2022). Essa linha de preocupação é sublinhada por Luciano Floridi, quando refere: “A AI não é meramente outra utilidade que precisa ser regulada quando estiver madura. É uma força poderosa que está remodelando nossas vidas, interações e nossos ambientes” (Floridi, 2022). Pela análise do cenário atual do desenvolvimento da IA, se poderá dizer que ela está apenas em sua “infância”, ou seja, ainda existe espaço para avanços e não se poderá duvidar que essa referida remodelação de Floridi poderá se concretizar.

A garantia do “controle do ser humano” assegura a proteção dos direitos e garantias fundamentais (no plano regulatório interno), além de ser um elemento protetor para salvaguardar os Direitos Humanos (no plano regulatório externo). Se entende o seguinte percurso da tradição: os Direitos Naturais representaram o estágio inicial que preparou a emergência dos Direitos Humanos, que geraram, na sequência da caminhada histórica da humanidade, os Direitos Fundamentais, a partir da inserção dessa amálgama da tradição nos textos constitucionais. Aprofundando e qualificando ainda mais essa trajetória, se poderá dizer que os Direitos de Personalidade representam a versão contemporânea desses movimentos históricos de lutas e sofrimentos, assegurando a sua previsão no principal texto legal do Direito Privado, o Código Civil (Engelmann, 2023).

Para se desenhar um projeto regulatório adequado para o “mundo da IA”, caberá trazer algumas categorias filosóficas: uma delas, é o *exercício da prudência*, da medida e da cuidadosa avaliação das diversas facetas envolvidas na formulação das respostas jurídicas. Elas deverão ser perspectivadas com o manejo das características que compõe o *phrónimos*:

[...] pensa-se que é característico de uma pessoa que tenha *phrónesis* ser capaz de deliberar bem acerca do que é bom e conveniente para si mesma, não em relação a um aspecto particular [...] e sim acerca das espécies de coisas que nos levam a viver bem de um modo geral (Aristóteles, 2001, 1140a).

Para que o *phrónimos* possa alcançar a realização adequada, ou a melhor possível, dos bens humanos básicos (conhecimento, vida, sociabilidade, dentre outros definidos por John Finnis (2011) exigirá o chamado *plano de vida racional* que busca atingir e atender às características genéricas dos desejos e necessidades humanas e seu ciclo de ocorrência.

Esses aspectos, relativos a um plano de vida, inspiraram John Finnis na formulação de uma das exigências básicas da razoabilidade prática, que foi denominada de *plano de vida coerente*. Cada ser humano, e por relação direta – cada sociedade – deve ter um conjunto de propósitos e orientações, organizados com determinado grau de harmonia, passíveis de serem concretizados. Não se trata de planos imaginários e sem possibilidade de realização, mas de compromissos efetivos (Finnis, 2011, p. 103). A vida do ser humano deve estar norteadada por um projeto, alicerçado nas vivências pessoais passadas e presentes, mas lançando as vistas para o futuro. Vale dizer, é irracional que o ser humano apenas pense no imediato, esquecendo-se de amarrar cada momento, como um contínuo caminhar de cada um no contexto social.

O referido plano de vida, apesar de sua pensada coerência, poderá ser objeto de reformas ou revisões, onde se abandonam velhos projetos e se elegem novos. Os bens humanos estão integrados e interagem com o plano de vida desenvolvido por cada ser humano, justificando, portanto, alterações no decorrer de sua execução ou a partir do momento em que um novo plano é traçado, especialmente para atender a algum bem humano básico que necessite, naquele momento, uma maior atenção.

O *plano de vida racional* implica responsabilidade do ser humano, que deverá considerar-se tanto no momento de estipular as metas, como na projeção dos resultados das ações desencadeadas. Valendo também para a regulação da IA. O espaço público e o espaço privado se conectam a partir da introdução do conceito de *phrónesis*, que representa a tomada de

decisões não voltadas somente ao privado. Nesse particular, emerge um aspecto característico: como as pessoas não são iguais, os planos eleitos por cada uma também sofrem variações. Isso destaca a pessoalidade das escolhas, vinculadas a uma série de contornos individuais, marcados pela chamada experiência de vida, e que pode ser chamada de pré-compreensão, profundamente influenciada pela tradição onde o sujeito está imerso. A razoabilidade introduzida nesse contexto indica que o *plano de vida coerente* não está ligado apenas a uma concepção individual. Pelo contrário, deverá ser traçado com vistas ao coletivo, do qual o individual faz parte, mediante a projeção da concepção aristotélica de *phrónimos*.

O *plano de vida coerente*, desenhado a partir da tradição, e em observância aos contornos apresentados pelo mundo da vida, sugere um modelo de ação, para orientar a tomada de decisões sobre os possíveis riscos trazidos pela IA. Por conta disso, a formulação de um *plano de vida coerente* para o grupo social deverá ser agregada a uma outra *exigência básica da razoabilidade prática*, que John Finnis denomina de *bem comum*, justamente para reforçar a necessidade de planos coerentes para a vida do ser humano na sociedade, onde se conjugue o individual com o coletivo, o público com o privado, sem a primazia, inicial, de nenhum deles. O bem comum

se refere ao fator ou conjunto de fatores que, presentes como considerações no raciocínio prático de uma pessoa, atribui *um sentido a*, ou *uma razão para*, sua colaboração com os outros, desde o seu ponto de vista, estruturando uma razão para sua colaboração entre eles mesmos e com essa pessoa” (Finnis, 2011, p. 154).

Esse conjunto de elementos poderá auxiliar na construção do caminho rumo à IA responsável.

4. Os princípios jurídicos como estruturas normativas para desenvolver a chamada IA responsável

Uma das alternativas que se pode apresentar para a regulação da IA, orientada pela atitude do *phrónimos* e organizada

a partir de um *plano coerente de vida*, poderá se dar por meio de modelos regulatórios estruturados a partir de princípios, que se encontram vinculados diretamente com a referida amálgama da tradição que inicia com os Direitos Naturais, se conectam aos Direitos Humanos e desembocam nas Constituições contemporâneas, a partir do capítulo dos Direitos Fundamentais. Como se desenvolverá essa perspectiva? Na revisão de literatura sobre o tema se encontraram diversas fontes, dentre as quais se destacam Weng e Izumo (2019), Zhang et al (2022) e *AI Ethics Guidelines Global Inventory* (2020). Nessa última fonte, foram sistematizados mais de 160 documentos, estruturados a partir de princípios, e que foram elaborados por atores públicos e atores privados, localizados em países do Sul, do Norte, do Leste e de Oeste global. Essas fontes destacam o papel regulatório dos princípios.

Os princípios integram a categoria geral de “normas jurídicas”, ao lado das regras. Portanto, os princípios também apresentam uma carga de “dever ser”, com potencial diferente daquele carregado pelas regras. O trabalho jurídico, na resolução de problemas, com o uso dos princípios, que são considerados “mandados de otimização”, emprega critérios hermenêuticos distintos daqueles que norteiam a aplicação das regras, que são “mandados definitivos”. A perspectiva de “atribuição de sentido hermenêutico” exige um rigor metodológico, pois é orientada pelas consequências que a movimentação principiológica busca atingir (Engelmann, 2001; Engelmann, 2023). Se vislumbra um espaço crescente dos princípios e uma perspectiva hermenêutica renovada na sua aplicação, especialmente na construção de modelos normativos para regular os avanços da IA. Os princípios apresentam uma linguagem aberta para se amoldar a novas perspectivas sociais e que exigem regulação ágil e flexível. Tais não são as características da regulação legislativo-estatal. Por isso, se considera o alargamento do campo de aplicação dos princípios: não apenas os princípios expressa e implicitamente formulados nas Constituições ou em outros documentos normativos, mas também os princípios que orientam o intérprete, seguindo as linhas axiológicas fundamentais do ordena-

mento jurídico, na formulação da regra e na justificativa da solução a ser dada ao caso (Zaccaria, 2022). Se pode destacar que nesse trabalho de orientação do intérprete se conjugam os princípios inseridos no texto constitucional, assim como os princípios relativos aos Direitos naturais-humanos, abrindo a produção normativa para a conjugação de elementos internos com os elementos externos ao ordenamento jurídico de cada Estado.

Os princípios permitem introduzir no mundo jurídico uma série de regras que vêm da economia e das ciências, mas também da moral e outros campos do saber, sublinhando uma capacidade “normogenética ilimitada”; e, como tal, desempenham uma função essencial na difícil governança do pluralismo das fontes do Direito e da complexidade sociojurídica trazida pela IA devendo ser redefinida e recalibrada argumentativamente caso a caso pelo intérprete (Zaccaria, 2022, pp. 20-26).

O que sem dúvida se perde em termos de certeza e determinação do Direito, nessa revalorização dos princípios, ganha-se em troca de maior pluralismo: não mais apenas os legisladores, mas também (e sobretudo) os juízes, não mais apenas os Estados, mas também os novos sujeitos particulares e novas agências reguladoras; não mais apenas a legislação estadual, mas também a legislação supraestatal, internacional, comunitária, global, contribuem para configurar o conjunto do Direito (Zaccaria, 2022, pp. 26-29). Esse o cenário criativo, poroso e flexível, onde os princípios jurídicos atuam na atualidade, promovendo a criação normativa de estruturas típicas de *autorregulação*, em alguns momentos; em outros, de *autorregulação regulada* (Engelmann, 2022).

Esse é o percurso do processo atual de *mutação do Direito*, já não centrada mais preponderantemente na lei, mas nos dois polos da Constituição e dos documentos internacionais dos Direitos Humanos, por um lado; e dos tribunais nacionais e internacionais de Direitos Humanos, por outro. Isso gera, segundo Zaccaria (2022), um quadro de porosidade do Direito, que acentua dramaticamente a sua contingência e indeterminação e enfraquece a sua capacidade de ordenar e dar forma

à convivência civil. Com isso, surge naturalmente o delicado problema de distinguir o que é Direito, do que não é Direito, e para diferenciar novas formas de Direito de formas espúrias e degeneradas (Zaccaria, 2022, pp. 28-29).

Com esse panorama, se pode desenhar a revalorização normativa dos princípios, que conduzirão o referido processo de distinção, que tem a arquitetura projetada nas redes digitais, em um panorama específico e inédito da era da hiperconectividade (Floridi, 2015), que se caracteriza pela indiferença de se viver no modo on-line ou no modo off-line, pois as relações humanas e as demais dela decorrentes ocorrem no modo *OnLife*, ou seja, elas se desenvolvem espaço especial denominado *Infosfera*, onde tudo poderá ser perfeitamente analógico e digital, off-line e on-line (Floridi, 2018). Essa nova perspectiva que exige a mutação *do* e *no* Direito, inovando-se a produção e o reconhecimento do que é jurídico, daquilo que não é albergado pelo Direito, notadamente pelo reconhecimento de procedimentos regulatórios que levam a produção dos efeitos jurídicos para além do legalismo que sempre assume a forma de uma proposta de modelos rígidos, que garantem os bens da certeza e da previsibilidade, mas intrinsecamente inadequados para os tempos instáveis em que se vive (Grossi, 2020, pp. 79-85), gerados pela convergência tecnológica, dentre as quais, uma delas é a IA.

A referida revalorização normativa dos princípios se liga ao percurso histórico iniciado pelos Direitos Naturais, passando pelos Direitos Humanos até se chegar aos Direitos Fundamentais. Essa “caminhada” evidencia o papel da historicidade que se deverá projetar no “acontecer histórico” orientado no movimento do círculo hermenêutico (Gadamer). Essa perspectiva se encontra alicerçada na perspectiva trazida por Paolo Grossi (2020, pp. 86-87): “a historicidade significa plasticidade, vontade de se conformar por um devir constante, sem ter a pretensão de interrompê-lo, fixando-o em uma espécie de modelo atemporal e, portanto, anti-histórico”. Valorizar a historicidade é sublinhar o papel da tradição no *design* da atribuição de sentido hermenêutico ao *acontecer da linguagem* que compõe o ser humano. Vale dizer,

[...] Historicidade significa basicamente humildade diante da complexidade redescoberta, singeleza diante da mudança e do novo que ela comporta; historicidade significa consciência de que, se o resultado de um Direito autenticamente humano deve ser alcançado, ele não pode ser reduzido a uma vontade potestativa, imposta de cima, mas deve ser remetido à busca dos muitos recursos que o fluxo histórico traz consigo: variando e confirmando, ou seja, ressignificando o percurso (Grossi, 2020, pp. 86-98).

O percurso que liga os Direitos Naturais aos Direitos Fundamentais se enriquece dessa historicidade e reafirma a sua importância contemporânea, ao sustentar o plano de vida coerente da sociedade, na busca da regulação principiológica, iluminada pelos elementos estruturantes da *phrónesis*.

Se o Direito é fruto da invenção, então o cânone da historicidade do Direito, como abordagem inquestionavelmente afável da dimensão jurídica, representa a bússola orientadora para o legislador e para a comunidade de juristas/intérpretes, bem como o resgate de um Direito que é sempre e apenas experiência, ou seja, uma dimensão da vida (Grossi, 2020, p. 98). Os princípios representam essa conjugação da experiência de vida da sociedade. Por isso, eles estão em condições de, ao se renovarem, servirem para se estruturar modelos de *autorregulação regulada* para as novidades e os desafios gerados a partir da IA.

Colhendo-se um exemplo organizado pelo Fórum Econômico Mundial, que compara os princípios destacados no documento produzido pelo *Chinese National New Generation Artificial Intelligence Governance Professional Committee* (CNNGAI-GPC) e o documento elaborado pelo *European High-level Expert Group on AI* (HLEGAI), se poderá ter uma ideia das distinções existentes, mas, paralelamente, as convergências principiológicas, que confirmam o papel que os princípios representam na arquitetura da governança regulatória criativa para a IA. Esses aspectos evidenciam a possibilidade de se fazer um diálogo regulatório global a partir de uma fonte do Direito que deverá ser revalorizada: os princípios.

Quadro 1: Comparativo de princípios entre China e União Europeia

Princípios Éticos propostos pelo Chinese National New Generation Artificial Intelligence Governance Professional Committee	Exigências chave formulados pelo European High-level Expert Group on AI
1. Harmonia e amizade	1. Bem-estar social e ambiental
2. Equidade e Justiça	2. Diversidade, não-discriminação e equidade
3. Tolerância e compartilhamento	3. Supervisão humana
4. Respeito à privacidade	4. Privacidade e governança de dados
5. Segurança e controlabilidade dos sistemas de IA	5. Robustez técnica e segurança dos sistemas de IA
6. Responsabilidades compartilhadas	6. Transparência
7. Colaboração aberta	7. Responsabilidade
8. Governança ágil	

Fonte: World Economic Forum (2021).

Em muitos aspectos, os princípios chineses parecem semelhantes aos da União Europeia, promovendo justiça, robustez, privacidade, segurança e transparência. As perspectivas metodológicas, no entanto, revelam diferenças culturais claras. De qualquer modo, o percurso regulatório por meio dos princípios evidencia sinais de aproximação e diálogo.

As diretrizes chinesas derivam de uma perspectiva focada na comunidade e orientada a objetivos. “Um alto senso de responsabilidade social e autodisciplina” é esperado dos indivíduos para participar harmoniosamente em uma comunidade que promove tolerância, responsabilidades compartilhadas e colaboração aberta. Essa ênfase é claramente informada pelo valor confucionista de “harmonia”, como um equilíbrio ideal a ser alcançado através do controle de paixões extremas – os conflitos devem ser evitados. Além de uma advertência severa contra “uso ilegal de dados pessoais”, há pouco espaço para regulação estatal. A regulação não é o objetivo desses princípios, que são concebidos para orientar os desenvolvedores da IA da “maneira certa” para a elevação coletiva

da sociedade (World Economic Forum, 2021). Aqui se observa o caráter histórico da sociedade chinesa, que influencia a sua concepção sobre o papel dos princípios, ratificando o seu caráter de “mandados de otimização”, com as consequências jurídico-sociais estruturadas em cada caso concreto.

Os princípios europeus, emergindo de uma abordagem mais focada em direitos e baseados em direitos, expressam uma ambição diferente, enraizada no Iluminismo e coloridos pela história da Europa. Seu objetivo principal é proteger os indivíduos contra danos bem identificados. Enquanto os princípios chineses enfatizam a promoção de boas práticas, a UE se concentra na prevenção de consequências malignas. Fica bem clara a concepção entre meios e fins que os princípios deverão atender. A orientação chinesa atrai uma direção para o desenvolvimento da IA, de modo que contribui para a melhoria da sociedade, o modelo europeu define as limitações aos seus usos, para que não aconteça às custas de certas pessoas (World Economic Forum, 2021). A regulação estatal acaba trazendo elementos nacionais que dificilmente se comunicarão com a regulação de outros Estados. Por isso, a estrutura regulatória formada a partir dos princípios poderá gerar redes de normatização que aproximam os Estados, em uma verdadeira concepção global da IA responsável, ou seja, preocupada em compatibilizar o desenvolvimento da IA, mas com a centralidade da proteção do ser humano.

5. Considerações finais

O tema da IA ganha importância gradativa, especialmente pelos desenvolvimentos de pesquisa nas áreas exatas, especialmente de Ciência da Computação e Computação Aplicada. Se tem uma demora na construção da regulação desses avanços. Por isso, o problema de pesquisa que orientou este capítulo de livro foi: “Como se deverá regular esses avanços, assegurando-se o desenvolvimento da chamada Inteligência Artificial responsável?”.

A regulação da IA deverá buscar metodologias que consigam equilibrar o desenvolvimento da pesquisa e dos pro-

cessos de inovação decorrentes, com a forma e o tempo dessa regulação. O processo legislativo tradicional não parece ser o caminho mais adequado, dada a demora das suas variadas etapas. Por isso, se apresentou a possibilidade de construção de modelos de “autorregulação regulada”, onde se mescla a iniciativa privada, como um “novo” ator para regular a IA, com a observância de elementos que se encontram em panorama externo a esses atores, como o caso dos Direitos Humanos. Esses foram aqui desenhados a partir das contribuições de John Finnis, notadamente pela orientação encontrada em um dos “requerimentos ou requisitos da razoabilidade prática”, que é a estrutura de *um plano de vida coerente*: que é considerado uma possibilidade de conjugação de esforços e pressupostos de interesse público com o interesse privado. Tal conjugação, orientada pela arquitetura aristotélica da *phrónesis*, conduzirá à escolha de princípios, globalmente aceitos, para se estruturar modelos de “autorregulação regulada”.

A aceitação normativa dos princípios vem encontrando respaldo em alguns autores da Teoria Geral do Direito, conforme se estudou no capítulo, e, em escala crescente, eles vêm surgindo em diversos documentos de órgãos públicos e privados, em um elevado número de países. Esse fenômeno mostra que os princípios poderão ser um denominador regulatório comum para a IA promovendo um diálogo global de autorregulação regulada.

Não se tem, ainda, uma confirmação definitiva dessa possibilidade de resposta ao problema antes proposto, mas se observa ser um caminho que vai ganhando a adesão crescente, justamente pela perspectiva porosa dos princípios. Eles se abrem para a transição regulatória, a partir da orientação proveniente da hermenêutica filosófica, de orientação de Martin Heidegger e Hans-Georg Gadamer. A caminhada foi iniciada. Existe um largo espaço para o avanço e aprofundamento dessa pesquisa, sem se esquecer, também, da possibilidade de se testar esses modelos em laboratórios reais, como é o caso do *Living Lab Regulatório*. Aqui se trouxe uma pequena construção de uma pesquisa maior que o autor está desenvolvendo.

Referências

- AI Ethics Guidelines Global Inventory* (2020) [Online]. Available at: <https://inventory.algorithmwatch.org/about> (Accessed: 5 March 2023).
- Aristóteles (2001) *Ética a Nicômaco*. Translated from Greek, introduction and notes by Mário da Gama Kury. 4th edn. Brasília: UnB, 1140a.
- Barbosa, M. M. (2021) *Inteligência artificial: entre a utopia e a distopia, alguns problemas jurídicos*. Coimbra: GestLegal.
- Codagnone, C., Liva, G. and Rodriguez de Las Heras Ballell, T. (2022) *Identification and assessment of existing and draft EU legislation in the digital field: Study for the special committee on Artificial Intelligence in a Digital Age (AIDA)* [Online], Policy Department for Economic, Scientific and Quality of Life Policies, European Parliament, Luxembourg. Available at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/703345/IPOL_STU\(2022\)703345_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/703345/IPOL_STU(2022)703345_EN.pdf) (Accessed: 6 March 2023).
- Cuervo Sánchez, C. A. (2021) Efectos de la inteligencia artificial en las estrategias de marketing: revisión de literatura. *Revista Internacional de Investigación en Comunicación aDResearch ESIC* (Primer semestre), 24(24), pp. 26-41 [online]. Available at: <https://doi.org/10.7263/adresic-024-02> (Accessed: 27 February 2023).
- Devang, V., Chintan, S., Gunjan, T. and Krupa, R. (2019) ‘Applications of artificial intelligence in marketing’. *Annals of the University Dunarea de Jos of Galati: Fascicle: XVII, Medicine*, 25(1), pp. 28-36, Galați, Romania.
- Engelmann, W. (2023) *Direito natural, ética e hermenêutica*. 2nd edn. revised, updated and expanded. Rio de Janeiro: Lumen Juris.

- Engelmann, W. (2022) 'Percursos para inovar a teoria geral das fontes do Direito: modelos de autorregulação regulada para as nanotecnologias, *sandbox regulatório* e princípios. Constituição, Sistemas Sociais e Hermenêutica' in Teixeira, A. V., Streck, L. L. and Rocha, L. S. (eds.) *Anuário do Programa de Pós-Graduação em Direito da Unisinos*, No. 18 [E-Book]. Blumenau, SC: Dom Modesto.
- Engelmann, W. (2001) *Crítica ao positivismo jurídico: princípios, regras e o conceito de Direito*. Porto Alegre: Sérgio Antonio Fabris.
- Finnis, J. (2011) *Natural law & natural rights*. 2nd edn. Oxford: Oxford University Press.
- Floridi, L. (2023) 'AI as agency without intelligence: on ChatGPT, large language models and other generative models'. *Philosophy and Technology* [online]. Available at: https://www.academia.edu/97068509/AI_as_Agency_without_Intelligence_on_ChatGPT_large_language_models_and_other_generative_models (Accessed: 5 March 2023).
- Floridi, L. (2022) 'The European Legislation on AI: a brief analysis of its philosophical approach' in Mökander, J. and Ziosi M. (eds.). *The 2021 Yearbook of the digital ethics lab, Digital Ethics Lab Yearbook*. Switzerland: Springer Nature.
- Floridi, L. (ed.) (2015) *The Onlife Manifesto: being human in a hyperconnected era*. London: Springer Open.
- Floridi, L. (2018) 'Soft Ethics and the Governance of the Digital', *Philosophy & Technology*, 31, pp. 1-8.
- Grossi, P. (2020) *Oltre la legalità*. Bari, Italy: Laterza & Figli.
- McKinsey Technology Trends Outlook (2022) *Applied AI* [Online] Available at: <https://www.mckinsey.com/~media/mckinsey/business%20functions/mckinsey%20digital/our%20in%20sights/the%20top%20trends%20in%20tech%202022/McKinsey-Tech-Trends-Outlook-2022-Applied-AI.pdf> (Accessed: 13 March 2023).

Regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento Inteligência Artificial) e altera determinados atos Legislativos da União Europeia (2021), Bruxelas, 21.4 COM(2021) 206 final. 2021/0106 (COD) [online]. Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF (Accessed: 6 March 2023).

Savage, N. (2023) ‘Why artificial intelligence needs to understand consequences’, *Nature*, Outlook, 24 February [online]. Available at: <https://doi.org/10.1038/d41586-023-00577-1> (Accessed: 2 March 2023).

Shah, H. (2021) ‘COVID-19 recovery: Science isn’t enough to save us’. *Nature*, 591, p. 503 [online]. Available at: <https://doi.org/10.1038/d41586-021-00731-7> (Accessed: 27 February 2023).

Smith, A. (2018) ‘Franken-algorithms: the deadly consequences of unpredictable code’, *The Guardian*, 30 August [online]. Available at: <https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger> (Accessed: 27 February 2023).

Stokel-Walker, C. (2023) ‘Generative AI is coming for the lawyers’, *WIRED*, Business, 21 February [online]. Available at: <https://www.wired.com/story/chatgpt-generative-ai-is-coming-for-the-lawyers/> (Accessed: 3 March 2023).

Zaccaria, G. (2022) *Postdiritto: nuove fonti, nuove categorie*. Bologna: Mulino.

Zhang, D., Maslej, N., Brynjolfsson, E., Etchemendy, J., Lyons, T., Manyika, J., Ngo, H., Niebles, J. C., Sellitto, M., Sakhaee, E., Shoham, Y., Clark, J. and Perrault, R. (2022) *The AI Index 2022 Annual Report*, AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March [online]. Available at: https://aiindex.stanford.edu/wp-content/uploads/2022/03/2022-AI-Index-Report_Master.pdf (Accessed: 5 March 2023).

Weng, Y.-H. and Izumo, T. (2019) 'Natural Law and its implications for AI Governance', *Delphi* (Winter), 3, 10 December [online]. Available at: https://works.bepress.com/weng_yueh_hsuan/118/ (Accessed: 6 March 2023).

World Economic Forum, 2021. *Can China and Europe find common ground on AI ethics?* 19 November [online]. Available at: https://www.weforum.org/agenda/2021/11/can-china-and-europe-find-common-ground-on-ai-ethics/?utm_source=sfmc&utm_medium=email&utm_campaign=2762835_Agenda_weekly-26November2021&utm_term=&emailType=Agenda%20Weekly (Accessed: 8 March 2023).

INTELIGENCIA ARTIFICIAL Y REGULACIÓN DENTRO DEL MARCO CONSTITUCIONAL ARGENTINO: ALGUNAS REFLEXIONES SOBRE EL IMPACTO DE LAS NUEVAS TECNOLOGÍAS

Martín Julián Acevedo – Miño¹

DOI: 10.29327/5312711.1-2

Sumario: 1. Introducción. 2. Una regulación dentro del marco constitucional. 3. La amenaza. 4. Reflexión final. Referencias.

1. Introducción

Cuando analizamos la irrupción de la inteligencia artificial en el mundo del derecho es razonable que, en tanto fenómeno novedoso, nos genere más dudas que certezas. Máxime cuando algunas noticias nos ponen en alerta sobre su desembarco y las eventuales consecuencias de su llegada. Algunos países y organizaciones reconocidas han prohibido la utilización de algunas herramientas de inteligencia artificial (Infobae, 2023; Lavanguardia, 2023). Los periódicos del mundo se inundan de especulaciones respecto a las herramientas que pone a disposición de las personas este avance tecnológico y, en su mayoría, comienzan a plantear la necesidad de su regulación (Peralta, 2023). Claramente, como lo indicaba un reciente artículo, estamos en presencia de un tipo de tecnología disruptiva entendida ésta como innovaciones que llegan para sustituir un proceso, un producto o una tecnología que ya está establecida, originando una nueva forma de operación, ya sea para los consumidores, las organizaciones empresarias, o para ambos; tienen la cualidad de transformar por completo

1 Abogado (UCSF) y doctor en Ciencias Jurídicas (UCA). Profesor de Derecho Constitucional (UCA). Profesor de doctorado (UCSF). Diputado constituyente estadual (Entre Ríos, 2008). Conjuez del Superior Tribunal de Justicia de Entre Ríos (2020/2022). Secretario General de la Asociación Argentina de Derecho Constitucional. ORCID: <https://orcid.org/0000-0001-9855-7576>.

una solución y la reemplazan, porque tienen atributos superiores que permiten abrir nuevas posibilidades (Santarelli, 2022).

Hace pocos años, en un análisis sobre el impacto de internet, que constituyó otra innovación revolucionaria en su momento, acertadamente se observaba que deviene central una reflexión sobre la relación entre derecho y técnica, toda vez que las innovaciones de naturaleza tecnológica vienen a incidir directamente sobre derechos y sobre la libertad, incluso económica, de los ciudadanos, remodelando contenidos, garantías y límites, dejando emerger nuevas problemáticas en relación al antiguo vínculo entre libertad y autoridad (Otranto, 2015, p. 67).

En la doctrina argentina y en la jurisprudencia de la Corte Suprema, la atribución de reglamentar los derechos se denomina “poder de policía”. Dicha facultad – entendida en un sentido restringido (en cuestiones vinculadas a la moralidad, salubridad y seguridad) o amplio (que podría incluir la defensa y promoción de los intereses económicos de la colectividad) – debe ejercerse dentro de los límites contenidos en la Constitución Nacional². El rol que debe jugar el Estado, como emisor de normas jurídicas, pero también como responsable de hacer cumplir las leyes, no debe traspasar dichos límites, bajo riesgo de ser declarado inconstitucional por la judicatura. Ante cada situación que el Estado regula, aún más cuando se trata de situaciones disruptivas, es altamente probable que su adecuación con la Carta Magna nacional sea debatida y el caso que nos ocupa no escapará a dicha posibilidad.

2. Una regulación dentro del marco constitucional

En el último siglo y medio, el constitucionalismo se ha desarrollado en casi todo el mundo, promoviendo una forma de estado que tiene como notas distintivas: a) la introducción, sin precedente en la historia, de una eficaz limitación de la actividad gubernamental por medio del derecho, cuya elabo-

2 Constitución de la Nación Argentina, disponible en <http://servicios.infoleg.gob.ar/infolegInternet/anexos/0-4999/804/norma.htm>.

ración debía corresponder a los mismos ciudadanos a través de la obra de sus representantes en el Parlamento; y b) la determinación precisa de dichos límites por medio de Constituciones escritas, las que muy pronto y en la mayoría de los casos aparecieron como “rígidas”, es decir, modificables sólo con procedimientos más complejos y solemnes que los empleados en el mecanismo legislativo ordinario (Biscaretti di Ruffia, 2006, p. 114). La Constitución Argentina sigue el modelo de constitución rígida, de inspiración liberal, que admite la limitación de derechos por parte del Estado, a través de leyes en sentido formal y material, asegurando “los beneficios de la libertad”, como lo indica su Preámbulo. Este esquema debería ser el marco – en tanto ley vigente – y en él deberíamos encontrar las respuestas a la regulación de la inteligencia artificial.

Como bien se ha dicho, la Constitución es norma, es decir, es un mandato que pretende regular conductas hacia el futuro; pero no es únicamente eso sino un fenómeno con múltiples dimensiones que, en su conjunto, van a actuar como fundamento del ordenamiento jurídico de una nación y, por tanto, la legitimidad, la validez y la vigencia de todas las leyes, actos de gobierno y conductas de las personas dependerá de su conformidad con ella (Barrera Buteler, 2019, p. 42).

En lo que a la República Argentina refiere, el art. 31 contiene dos principios sustantivos del ordenamiento jurídico: el principio de supremacía constitucional y el principio de jerarquía de las fuentes del derecho. El art. 31 declara la supremacía de la Constitución Nacional por sobre todo el orden jurídico argentino. Éste debe subordinarse a aquélla y no debe modificarla si no es por el procedimiento de constitución rígida establecido en el art. 30 de la Ley Suprema, disposición que crea el poder constituyente derivado y el modo y alcance de su ejercicio. Cuadra recordar además que por disposición del art. 27 de la Constitución Nacional también los tratados de paz y comercio deben respetar el orden público establecido en la norma de base (Gelli, 2005, p. 364).

En consecuencia, el orden constitucional argentino admite regulaciones y por tanto admitiría también esos límites en la utilización de herramientas de inteligencia artificial, te-

niendo siempre en la mira el principio de supremacía constitucional. Pero, ¿cuál es el margen de maniobra para el establecimiento de una normativa limitante? Cuando de intervención del Estado se trata ante novedosas irrupciones de los adelantos tecnológicos, no podemos dejar de pensar en el precedente del “Cine Callao” (CSJN, Callao (Cine) s/ interpone recurso jerárquico c/ resolución dictada por la Dirección Nacional del Servicio de Empleo, 1960) de la Corte Suprema de la Nación. Allí la máxima autoridad judicial argentina – hace más de medio siglo – debió expedirse sobre la regulación atinente a espectáculos cinematográficos y a la obligación de incluir números en vivo de los empresarios del sector, atento el desempleo que implicaba el fenómeno del cine entre los artistas de teatro. En dicha oportunidad, el Tribunal dijo que, entendiendo el poder de policía en sentido amplio, consideraba “legítimamente incluida la facultad de sancionar disposiciones legales encaminadas a prevenir, impedir, morigerar o contrarrestar, en forma permanente o transitoria, los graves daños económicos y sociales susceptibles de ser originados por la desocupación en mediana o gran escala” (CSJN, Callao (Cine) s/ interpone recurso jerárquico c/ resolución dictada por la Dirección Nacional del Servicio de Empleo, 1960), entendiendo apropiada y ajustada a la Constitución la normativa atacada.

En otro ya legendario pronunciamiento, el máximo Tribunal recordó que “Ni el derecho de usar y disponer de la propiedad, ni ningún otro derecho reconocido por la Constitución, reviste el carácter de absoluto. Un derecho ilimitado sería una concepción antisocial. La reglamentación o limitación del ejercicio de los derechos individuales es una necesidad derivada de la convivencia social. Reglamentar un derecho es limitarlo, es hacerlo compatible con el derecho de los demás dentro de la comunidad y con los intereses superiores de esta última”. La argumentación fue contundente al recordar que “La misma Constitución ha consignado limitaciones especiales respecto de algunos derechos; pero no siendo posible prever ni establecer en ella todas las condiciones a que sería menester subordinarlos para hacerlos adaptables a la vida de relación, ha confiado al poder legislativo la misión de regla-

mentar su ejercicio, poniendo al mismo tiempo un límite a esa facultad reguladora (arts. 14 y 28)” (CSJN, Ercolano, Agustín c/ Lanteri de Renshaw, Julieta, 1922).

No escapa a nuestro conocimiento la prolífica jurisprudencia de la Corte respecto al ejercicio del poder de policía, pero quizás el citado precedente “Cine Callao” resulta apropiado para el fenómeno tecnológico que abordamos. Precisamente, porque fue ante una situación novedosa (la llegada del cinematógrafo) en la que la autoridad debió regular su impacto en la vida económica y social del país. Quizás algo similar deba ocurrir con las herramientas de la inteligencia artificial, toda vez que ella es uno de los vectores sobre los que se potencia el fenómeno tecnológico al que se asiste, involucrándose en todos los procesos innovadores (Santarelli, 2022).

3. La amenaza

En estudios recientes se ha afirmado, consistentemente, que estamos viviendo una época de “fascinación bigdataísta”, o de exaltación de la robótica y de la inteligencia artificial, apuntando que las decisiones que afectan a nuestras vidas no están hechas por humanos, sino por modelos matemáticos, hasta el punto de que se han hecho dueños de las actividades humanas y casi del saber humano, dada la capacidad de generar conocimiento propio, y tomar decisiones por sí mismos, mediante aprendizaje automático (*machine learning*) y aprendizaje profundo (*deep learning*) (Goñi Sein, 2019, p. 60).

Esa exaltación, por un lado, y un desarrollo que no registra antecedentes en la historia de la humanidad, por otro, lógicamente genera temores, absolutamente fundados, por cierto. La tarea de los juristas será la de despejar toda incertidumbre respecto a los posibles efectos nocivos de estas herramientas. La Comisión Europea ha remarcado que “como sucede con toda nueva tecnología, el uso de la IA presenta tanto oportunidades como amenazas”, agregando que “los ciudadanos temen quedarse indefensos a la hora de proteger sus derechos y su seguridad frente a los desequilibrios informativos de la toma de decisiones mediante algoritmos, y las empresas

sienten inquietud debido a la inseguridad jurídica” (Comisión Europea, 2020, p. 11). También la oficina del Alto Comisionado de Derechos Humanos de Naciones Unidas ha expresado que “mientras mayor sea el riesgo para los derechos humanos, más estrictos deben ser los requisitos legales para el uso de la tecnología de IA”. A ello agregó que “... dado que la evaluación de los peligros y cómo superarlos puede llevar todavía algún tiempo, los Estados deberían implementar desde ahora moratorias sobre el uso de las tecnologías que presentan un alto potencial de riesgo” (Alto Comisionado de las Naciones Unidas para los Derechos Humanos, 2021).

Por ello, desde que la IA puede ser descrita como un sistema basado en la decodificación de patrones de comportamiento contenidos en datos, pudiendo determinar regularidades que constituyen predictivos que la habilitan a ejecutar cambios en la realidad en la que opera, es posible predicar su autonomía; es decir, la capacidad de actuar por sí misma, sin control humano: lo que se traduce en que una vez diseñado el software inteligente, es capaz de actuar conforme a la predicción que es resultado del proceso (Santarelli, 2022), y ello se advierte como una amenaza en varios aspectos. Ante tal circunstancia, emerge la obligación del Estado de abordarlo con una regulación específica que contemple la supervisión y decisión humana, que tenga especialmente en cuenta el derecho a la intimidad y protección de datos, que no atente contra la dignidad humana, entre otros aspectos fundamentales, y que observe los límites de la razonabilidad, como principio ineludible.

Pero, acertadamente, un autor local remarca que el creciente y acelerado proceso del desarrollo tecnológico no se compadece con el lento y reflexivo avance de la ciencia jurídica, que en su doble faz dogmática-normativa tiene su propio proceso elaborativo, el cual se ve desbordado por las urgencias propias del primero, propio de las ciencias duras (Saux, 2022, p. 126). El desarrollo tecnológico impone tiempos que no son los habituales del mundo del derecho.

Ahora bien, ante la irrupción imparable a la que aparentemente asistimos, ¿tenderemos a esgrimir el “*Necessitas non*

habet legem” como reza un antiguo proverbio cuyo significado es “la necesidad no tiene ley”? ¿Cómo operará la costumbre como fuente del derecho en estos temas? Siempre que referimos a normas y potestades en tiempos de crisis sobrevuela ese concepto de que la necesidad no tiene ley, algunas veces en forma más explícita que en otras. ¿La necesidad de asimilar las nuevas tecnologías nos pondrá en escenarios sin ley?

4. Reflexión final

Coincidimos con la apreciación de que “de alguna manera la interrelación entre el Derecho y las llamadas nuevas tecnologías (informática, telemática, robótica, ingeniería genética, biomedicina, nanotecnología, entre otras) suscita para el jurista grandes desafíos y preocupaciones, en relación a lo cual se ha dicho, atinadamente, que el abordaje de la cuestión revela una característica singularísima, cual es “...el insuperable quiebre de la simetría entre lo fáctico y lo jurídico...” (Saux, 2022, p. 126). En los Evangelios existe una expresión que puede resultar aplicable al tema que analizamos. Según los textos sagrados, Jesús dijo “Nadie usa un pedazo de género nuevo para remendar un vestido viejo, porque el pedazo añadido tira del vestido y la rotura se hace más grande”, y también Cristo contundentemente afirmó que “Tampoco se pone vino nuevo en odres viejos, porque los odres revientan, el vino se derrama y los odres se pierden. ¡No, el vino nuevo se pone en odres nuevos, y así ambos se conservan!”³. Las enseñanzas evangélicas nos llevan a reflexionar respecto a la adecuación de lo existente, los odres viejos, y su capacidad para receptar las novedades tecnológicas, el vino nuevo.

¿Serán las leyes existentes suficientes para atender el fenómeno de las herramientas de inteligencia artificial? ¿Funcionará el sistema actual como un vestido viejo que intentamos remendar con un género nuevo? Todo hace pensar que el entramado jurídico existente no resultará apropiado y que se necesitarán nuevas leyes para regular lo atinente a las novedo-

3 Evangelio de San Marcos, 9:16-17, disponible en: https://www.vatican.va/archi ve/ESL0506/___PUJ.HTM.

sas herramientas tecnológicas. Vino nuevo en odres nuevos. Pero, ¿será el orden constitucional existente en los Estados capaz de contener el arrollador desembarco de herramientas tecnológicas nunca antes vistas? Está claro que – en gran medida – el lento y reflexivo avance de la ciencia jurídica responde en muchas situaciones al esquema constitucional vigente. ¿No será el orden constitucional también un odre viejo que “reventará” ante el vino nuevo de la inteligencia artificial?

Referencias

- Alto Comisionado de las Naciones Unidas para los Derechos Humanos (2021) *Los riesgos de la inteligencia artificial para la privacidad exigen medidas urgentes*. Ginebra: s.n.
- Barrera Buteler, G. (2019) *Derecho Constitucional*. Córdoba: Advocatus.
- Biscaretti di Ruffia, P. (2006) *Introducción al derecho constitucional comparado. Las “formas de Estado” y las “formas de gobierno”*. *Las Constituciones modernas y 1988-1990 un trienio de profundas transformaciones constitucionales en Occidente, en la URSS y en los estados socialistas del Este europeo*. s.l.:Fondo de Cultura Económica.
- Comisión Europea (2020) *Libro Blanco sobre la inteligencia artificial – un enfoque europeo orientado a la excelencia y la confianza*. s.l.:s.n.
- CSJN (1922) *Ercolano, Agustín c/ Lanteri de Renshaw, Julieta*. s.l.:s.n.
- CSJN (1960. *Callao (Cine) s/ interpone recurso jerárquico c/ resolución dictada por la Dirección Nacional del Servicio de Empleo*. s.l.:s.n.
- Gelli, M. A. (2005) *Constitución de la Nación Argentina comentada y concordada*. 3ra edición ed. Buenos Aires: La Ley.
- Goñi Sein, J. L. (2019) Innovaciones tecnológicas, inteligencia artificial y derechos humanos en el trabajo. *Documentación Laboral Revista de la Asociación Española de Derecho del Trabajo y la Seguridad Social*, Issue N°11, p. 60.

- Infobae, (2023) *Infobae*, 30 March [online]. Available at: <https://www.infobae.com/america/agencias/2023/03/30/de-nuncian-en-estados-unidos-el-chatgpt-de-openai-y-piden-que-sea-suspendido/> (Accessed: 4 September 2023).
- La Vanguardia (2023) 'Italia bloquea el uso de ChatGpt por no respetar la ley de protección de datos', *La Vanguardia* [online]. Available at: <https://www.lavanguardia.com/tecnologia/20230331/8868073/italia-prohibe-chatgpt-respetar-legislacion-datos.html> (Accessed: 4 September 2023).
- Otranto, P. (2015) *Internet nell'organizzazione amministrativa: reti di libertà*, Università degli studi di Bari "Aldo Moro". Bari: Cacucci Editore.
- Peralta, L. A. (2023) *El País*, 11 April [online]. Available at: <https://cincodias.elpais.com/economia/2023-04-11/es-realmente-necesaria-una-regulacion-de-la-inteligencia-artificial.html> (Accessed: 4 September 2023).
- Santarelli, F. G. (2022) *La madeja de la inteligencia artificial. En busca de la punta del hilo*. Buenos Aires: La Ley Thomson Reuters.
- Saux, E. (2022) 'Algunas digresiones sobre el impacto de la inteligencia artificial en el derecho' in Andruet, A. S. *Impactos y alcances de la Inteligencia Artificial en el Derecho y en el Derecho Judicial*. Ciudad Autónoma de Buenos Aires: La Ley; Córdoba: Academia Nacional de Derecho y Ciencias Sociales de Córdoba.

DISCRIMINAÇÃO ALGORÍTMICA E COLONIALISMO DIGITAL

José Luiz de Moura Faleiros Júnior¹

DOI: 10.29327/5312711.1-12

Resumo: Os efeitos do colonialismo persistem até os dias atuais, embora muitas de suas características tenham sido transformadas por meio de tecnologias de informação digital que propiciam ambientes baseados em algoritmos, nos quais a falta de regulamentação e governança perpetra a exploração de vulnerabilidades. Essas características coloniais são equivalentes às atuais atividades de informação digital e referem-se ao conceito mais geral de colonialidade. Assim, ao contrário do colonialismo, a colonialidade continua expressando formas de relacionar poder, saber, ser e gênero, mas em um ambiente mais sofisticado. O surgimento de novas Tecnologias de Informação e Comunicação, baseadas nos mecanismos da Inteligência Artificial, têm favorecido um ambiente virtual denominado Infosfera, no qual organismos informacionalmente incorporados (*inforgs*), bem como o fortalecimento de um colonialismo digital formado por empresas tecnológicas do Norte Global de modo a permitir o acúmulo de riquezas e lucros às custas das desigualdades e injustiças sociais impostas por um novo colonialismo hegemônico que emerge e é pautado pelo capitalismo de vigilância, e que, mais uma vez na história, provoca violência epistêmica em relação a outras fontes virtuais de conhecimento, na medida em que considera as pessoas o produto ou a matéria-prima desse sistema nefasto. O artigo tem como objetivo

1 Doutorando em Direito Civil pela Universidade de São Paulo – USP/Largo de São Francisco. Doutorando em Direito, na área de estudo “Direito, Tecnologia e Inovação”, pela Universidade Federal de Minas Gerais – UFMG. Mestre e Bacharel em Direito pela Universidade Federal de Uberlândia – UFU. Especialista em Direito Digital. Especialista em Direito Civil e Empresarial. Advogado e Professor. ORCID: <https://orcid.org/0000-0002-0192-2336>. E-mail: jfaleiros@usp.br.

principal o estudo do pensamento descolonial no já citado contexto do novo colonialismo digital hegemônico no Brasil. Para tanto, utilizar-se-á o método dedutivo, a partir da análise de premissas gerais para se chegar a uma determinada conclusão, bem como da pesquisa bibliográfica.

Palavras-chave: Colonialismo digital. Perfilização. Discriminação algorítmica. Pensamento descolonial. Direitos humanos.

Abstract: The effects of colonialism persist until the present day, although many of its characteristics have been transformed through digital information technologies that enable algorithmically-driven environments in which the lack of regulation and governance perpetrates the exploitation of vulnerabilities. These colonial characteristics are equivalent to current digital information activities and they refer to the more general concept of colonality. Thus, unlike colonialism, colonality continues as a way of relating power, knowledge, being and gender, but on a more sophisticated environment. The emergence of new information and communication technologies, based on the mechanisms of Artificial Intelligence, has favored a virtual environment called Infosphere, in which informationally embodied organisms (*inforgs*), as well as the strengthening of a digital colonialism formed by technological companies from the Global North, allow the accumulation of wealth and profit at the expense of inequalities and social injustices imposed by a new hegemonic digital colonialism that emerges and is guided by surveillance capitalism, which, once again in history, causes epistemic violence in relation to other virtual knowledge sources, insofar as it considers people as the product or raw material of this nefarious system. The article has its main objective focused on the study of decolonial thinking in the aforementioned context of new hegemonic digital colonialism in Brazil. To achieve that goal, the deductive approach method shall be used, from the analysis of general premises in order to reach a particular conclusion, as well as bibliographic research.

Keywords: Digital colonialism. Profiling. Algorithmic discrimination. Decolonial thinking. Human rights.

Sumário: 1. Introdução. 2. O colonialismo digital e os mercados ricos em dados. 3. A proteção aos direitos humanos e os limites à inovação desregulada. 4. Conclusão. Referências.

1. Introdução

O avanço acelerado nas técnicas de aprendizado de máquina (*machine learning*) e profundo (*deep learning*) está tornando cada vez mais comum a utilização de algoritmos para agilizar processos e melhorar resultados. Com este novo paradigma, muitas atividades econômicas já foram automatizadas e categorizam pessoas com base no que é chamado de *profiling*, devido ao manuseio de grandes quantidades de dados (*Big Data*).

Em paralelo ao desenvolvimento tecnológico, houve o surgimento de um verdadeiro colonialismo digital, exercido por empresas de tecnologia sediadas principalmente no norte global, que impõem suas visões do mundo ao sul, incluindo o Brasil, através de aplicativos, algoritmos e redes sociais. Para lidar com os questionamentos e proteger as vulnerabilidades neste universo baseado em dados, leis foram criadas ao redor do mundo. No entanto, as consequências do colonialismo ainda persistem, mesmo com a transformação de algumas de suas características através da tecnologia.

Essas características coloniais são similares às atividades atuais de informação digital e se referem ao conceito mais amplo de colonialidade, que pode ser avaliado do ponto de vista sociológico ou antropológico para a reconstrução do pluralismo jurídico no século XXI. A colonialidade revela conexões entre poder, conhecimento, ser e gênero em um ambiente mais avançado, em contraste ao colonialismo. O surgimento de novas Tecnologias de Informação e Comunicação, essencialmente baseadas no implemento de algoritmos de Inteligência Artificial, tem favorecido o incremento da Infosfera, na qual organismos informacionalmente corporificados (*inforgs*), bem como o fortalecimento de um colonialismo digital formado por empresas de tecnologia do Norte, propiciam o acúmulo de riquezas e lucros às custas das desigualdades e injustiças sociais impostas por um novo colonialismo digital hegemônico, que é pautado pelo capitalismo de vigilância.

O conceito de capitalismo de vigilância traz à tona traços de uma nova ordem econômica que considera a experiência humana fonte de informações para algoritmos de aplicação comercial. Geralmente, são baseados em dados que prenunciam a oferta de serviços supostamente gratuitos por empresas de tecnologia para as pessoas, a fim de monitorar seus comportamentos com um nível muito alto de detalhamento, uma vez que aquelas lucram assustadoramente com a apropriação de dados dessas, capturados por algoritmos e processados por técnicas desumanizantes de perfilização e estigmatização.

A ânsia pela aceleração econômica traz como consequência a precipitação quanto ao uso de certos algoritmos, os quais, embora possam ser programados para que tenham funcionamento otimizado, ao serem alimentados por quantidades colossais de dados (a denotar a razão do emprego da expressão *Big Data*), apresentam conclusões baseadas em estatísticas que, no atual estado do desenvolvimento tecnológico, ainda são operacionalizadas por estruturas semânticas. Dependem, ainda que a nível revisional, da participação de um indivíduo humano – embora isso nem sempre seja observado, ou mesmo fiscalizável – e, não por outra razão, o artigo 20 da Lei Geral de Proteção de Dados Pessoais (Lei nº 13.709/2018) prevê o direito à revisão das decisões automatizadas, realizadas exclusivamente por algoritmos.

E, para muito além da preocupação com os riscos inerentes a essa temática, o que já se tem são algoritmos falhos, seja porque foram mal concebidos por seus desenvolvedores, seja porque foram incrementados a partir de bases de dados inadequadas, incompletas ou mal alimentadas, seja porque falham em reconhecer e valorar identidades culturais que, durante séculos, foram ignoradas pelo colonialismo eurocêntrico e ocidental.

Escritos doutrinários recentes indicam como essas situações punem pessoas em razão de estruturas técnicas alimentadas por dados coletados de inúmeras fontes e que abastecem algoritmos hipercomplexos. Há casos em que sequer seus desenvolvedores compreendem totalmente seu funcionamento e isso desencadeia um problema grave do ponto de

vista regulatório e da imperiosa necessidade de atendimento aos deveres de proteção assumidos pelo Estado.

O conceito de capitalismo de vigilância, densamente explorado nos escritos de Shoshana Zuboff – marco teórico da pesquisa – traz à tona os traços de uma nova estruturação dos mercados nos quais a autora considera a experiência humana fonte de informações e alimentação de algoritmos de aplicação comercial usualmente sutil (e muitas vezes imperceptível), geralmente voltados à oferta de serviços supostamente gratuitos para pessoas que sequer sabem quais são os riscos envolvidos nessas práticas.

O pensamento descolonial é definitivamente um assunto de grande relevância, e discutir criticamente a realidade sempre desafiadora das poucas *Big Techs* que, atualmente, dominam a Infosfera é um passo importante em direção à abordagem ética desejada no desenvolvimento da IA brasileira, tendo em vista que, embora não se possa vislumbrar, antecipadamente, todas as consequências do implemento da técnica na sociedade da informação, ao menos é possível compreender alguns de seus problemas para que eventual mapeamento de riscos propicie a parametrização de estruturas de proteção que congreguem inovação e regulação em esperada harmonia, potencializada pelo pensamento descolonial como forma de rompimento do novo colonialismo digital de Inteligência Artificial imposto por pequenos grupos de empresa de tecnologia.

O artigo tem como objetivo principal o estudo do pensamento descolonial no já citado contexto do novo colonialismo digital hegemônico no Brasil. Para tanto, deve-se utilizar o método dedutivo, a partir da análise de premissas gerais para se chegar a uma determinada conclusão, bem como da pesquisa bibliográfica.

2. O colonialismo digital e os mercados ricos em dados

O advento da chamada *web 3.0* marcou o atual estágio da sociedade da informação (Hayles, 2012, p. 55), que avança a passos largos para a proliferação da Internet das Coisas (*Internet of Things*, ou *IoT*), que também se situa na fronteira

da chamada *web 4.0* ou Internet de Todas as Coisas (*Internet of Everything*, ou *IoE*), particularmente no contexto da propagação de novas estruturas de telefonia móvel pela tecnologia 5G (Gajewski and Krawiec, 2016, pp. 275 e seg.). Em síntese, tudo passa a ser funcionalizado por estruturas preditivas baseadas em dados.

Com o fluxo incessante desses dados, preocupações emergem quanto aos riscos da hiperconectividade (Greenard, 2015, p. 58), uma vez que “a IoT pode ser vista em diferentes dimensões pelos diferentes setores da academia e da indústria; qualquer que seja o ponto de vista, a IoT ainda não atingiu a maturidade e é vulnerável a todos os tipos de ameaças e ataques.” (Peppet, 2014, pp. 147 e seg.) A ânsia pela aceleração econômica tem como consequência hodierna a precipitação quanto ao uso de certos algoritmos, os quais, embora possam ser programados para que tenham funcionamento otimizado (e desejável alto índice de acerto em suas previsões), ao serem alimentados por quantidades colossais de dados (*Big Data*), apresentam inferências baseadas em estatísticas que, no atual estado da arte, ainda são operacionalizadas por estruturas semânticas (Domingos, 2015, p. 6).

O estudo da Inteligência Artificial está intrinsecamente ligado à amplitude do valor atribuído aos dados na sociedade da informação e ao que Yuk Hui (2019), designa como recursividade nos usos determinísticos propiciados pelo uso da técnica para fins de controle.² A aglutinação de dados propicia a formação de grandes acervos (*Big Data*)³ que fomentam o hodierno debate em torno da pujante e empolgante revolução informacional (Floridi, 2010, p. 3 e seg.), propulsão pela evolução da computação, alimentada e robustecida por acervos informacionais antes inimagináveis.

2 Anota o autor: “Control through tertiary retentions and protensions such as surveillance, social credits, and big data analysis is taking the first path, in which recursive machines are integrating individuals as the constituents of computation. What Deleuze calls the society of control is fully demonstrated in our digital epoch, of which digital control and flexibility (e.g., modulation or performativity) are its means. We may want to say that it is a mechanist use of organicist machines for deterministic use” (Hui, 2019).

3 Em simples termos, eis o conceito: “Big Data is all about seeing and understanding the relations within and among pieces of information that, until very recently, we struggled to fully grasp”. (Mayer-Schönberger, 2014, p. 19).

Insofismavelmente, o caminho que se está a trilhar é marcado pela empolgação excessiva em torno da automação e da aceleração de processos, e pela otimização algorítmica de resultados lucrativos.⁴ Escritos doutrinários recentes indicam como essas situações punem pessoas em razão de modelos preditivos alimentados por dados coletados de inúmeras fontes e que abastecem algoritmos hipercomplexos (Muntadas, 2021, pp. 641-658). De certo modo, o *machine learning* lhes permite produzir autossofisticação tão acelerada que, a partir de determinado momento, perde-se a possibilidade de rastrear suas mudanças a ponto de sequer seus desenvolvedores compreenderem totalmente seu funcionamento.

Isso desencadeia um problema grave do ponto de vista regulatório e da imperiosa necessidade de atendimento aos deveres de proteção assumidos pelo Estado, pois novos sistemas passam a ser estruturados ao alvedrio da participação estatal, conjugando normas e sanções a um só tempo, o que remete à própria gênese da discussão sobre o pluralismo jurídico que marcou a passagem do século XIX para o século XX (Rouland, 2003, pp. 182-184).

Se, no passado, já se enunciava a ideia de que vários direitos poderiam interagir no contexto interno de uma mesma sociedade, o que denotava a preponderância do aspecto territorial dessa aferição, atualmente, as estruturas de mercado se tornam cada vez mais dependentes da atenção, pois dados são coletados para produzir as mais estratificadas soluções baseadas em preferências dos usuários, formando ecossistemas baseados em controle⁵, nos quais os algoritmos ‘mais eficazes’ e mais desejados são os que coletam e estratificam mais dados – gerando desigualdades⁶ –, os processam em

4 Importante o comentário de Jaron Lanier: “*Today, we can still think of information as the intangible enabler of communications, media, and software. But as technology advances in this century, our present intuition about the nature of information will be remembered as narrow and shortsighted. We can think of information narrowly only because sectors like manufacturing, energy, health care, and transportation aren’t yet particularly automated or ‘net-centric’*” (Lanier, 2013, p. 22).

5 Zuboff (2019a, p. 4) diz: “*Entanglements of knowledge, authority and power are no longer confined to workplaces as they were in the 1980s. Now their roots run deep through the necessities of daily life, mediating nearly every form of social participation*”.

6 Eubanks, V. (2018) *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin’s Press; O’Neil, C. (2016) *Weapons of Math*

menor tempo e oferecem as mais rápidas respostas à finalidade operacional para a qual foram desenvolvidos, impondo novos desafios ao direito⁷, que se depara com a premente necessidade de “descolonizar o pensamento, pensar desde a fronteira, propor um paradigma outro ou desobediência epistêmica” (Zuboff, 2019a, p. 4).⁸

O surgimento de marcos regulatórios voltados à proteção de dados pessoais em todo o mundo ainda é um fenômeno recente, mas claramente insuficiente para se contrapor ao problema em questão. Isso porque a melhoria das informações e da ética da IA para o desenvolvimento de algoritmos têm sido amplamente discutidas e elogiadas como meios adequados para mitigar os impactos de tais práticas indesejadas, embora o desafio seja enorme (Eubanks, 2018; O’Neil, 2016). Embora isso ainda possa ser considerado por alguns como uma ideia rebuscada em um mundo de variedade intercultural, abordagens pluralistas para sua compreensão são uma forma essencial de melhorar a regulamentação para o desenvolvimento de algoritmos em um contexto marcado por conhecimentos epistêmicos eurocêntricos⁹ que marcam o conceito de colonialidade (Mignolo, 2017).

É um novo ecossistema, virtualizado (eis que realizado no ciberespaço), com regras próprias¹⁰ que passam a reger

Destruction: How Big Data Increases Inequality and Threatens Democracy. New York: Crown.

- 7 Conferir, por exemplo, Thatcher, J., O’Sullivan, D. and Mahmoudi, D. (2016) Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data. *Environment and Planning D: Society and Space*, Preprint, 1-17. doi: 10.1177/0263775816633195. (Accessed: 10 January 2023).
- 8 Bragato, F. F. (2014) Para além do discurso eurocêntrico dos direitos humanos: contribuições da descolonialidade. *Revista Novos Estudos Jurídicos*, 19(1), p. 214.
- 9 Cave and Dihal (2020, p. 700) anotam: “*This process is part of what can be described as decolonising AI: a process of breaking down the systems of oppression that arose with colonialism and have led to present injustices that AI threatens to perpetuate and exacerbate.*” Conferir, ademais, as considerações de: Quijano, A. (2005) *Colonialidade do Poder. A Colonialidade do Saber: Eurocentrismo e Ciências Sociais. Perspectivas Latino-americanas*. Buenos Aires: CLACSO.
- 10 Se reportando ao autor polonês Leon Petrażycki, Norbert Rouland destaca que o autor “eleva à dignidade do direito os sistemas de normas que concorrem com o direito oficial (regras de jogos, códigos esportivos, leis do meio, jogos infantis etc.)” (Rouland, 2003, p. 184). Em perspectiva mais radical, valiosa a lembrança da proposta de Jacques Vanderlinden. O autor constrói uma visão de pluralismo jurídico baseada em 4 eixos: a) o direito somente pode ser concebido no contexto de uma sociedade; b) o direito consiste de mecanismos; c) o direito não

os próprios sentidos das sociedades. De fato, a partir dessa transformação hodierna, se torna possível reconhecer que a pluralidade dos ordenamentos não repousa apenas na multiplicidade dos grupos ou associações. Junto ao direito social, manifesta-se o direito interindividual, que é baseado nos intercâmbios entre indivíduos. A clivagem formal/informal na realidade social do direito não fornece uma representação adequada da relação entre o direito do Estado e o direito da sociedade, pois cada há uma dinâmica subjacente entre o direito organizado e o “não organizado” que não pode ser ignorada.¹¹

Para que não haja imprecisão conceitual, é importante destacar que o termo colonialidade – mais abstrato e geral – sucede e sobrevive ao conceito de colonialismo, cuja marca essencial é a “relação política e econômica de poder e dominação colonial de um povo, Estado ou não sobre o outro” (Zeifert and Agnoletto, 2019, pp. 197-218). Noutros termos, a colonialidade pode ser considerada aspecto essencial do poder inserido nas relações de dominação coloniais modernas, ao passo que o colonialismo é um “processo de poder” (Bragato, 2014, p. 212).

Com efeito:

Nos países pobres ou de renda média e regiões desprovidas de capital e sem núcleos de desenvolvimento tecnológico, o processo de extração de dados equivale a perda de riqueza, de valor gerado pelas subjetividades e intersubjetividades dessas sociedades. Submetidas à doutrina neoliberal e com a mentalidade do colono rico, as elites econômicas e grande parte dos líderes políticos desses países tentam impedir a construção de alternativas à extração de dados pelas plataformas, bloqueiam a formulação de políticas pela autonomia, pela proteção de dados e tecnopolíticas que enfrentem o paradigma dominante (Silveira, 2020, p. 180).

necessariamente surge como sistema; d) o pluralismo jurídico diz respeito à existência de diferentes mecanismos aplicáveis a situações idênticas dentro de uma mesma ordem social (Vanderlinden, 1989, pp. 149-157).

- 11 Nesse aspecto, importante a menção ao conceito de ‘infra-direito’, que lastreia a teoria do pluralismo jurídico proposta por Arnaud, que já era reconhecida por Jean Carbonnier, e que passa a nortear a reformulação da visão que se tem do pluralismo jurídico. Esse tema pode ser conectado à ideia de “campo jurídico vulgar” (Arnaud) em que o direito se contrasta frequentemente com “práticas vividas” e “visões concebidas”. Sobre o tema, cf. Villas Bôas Filho, O. (2019) Uma abordagem sociológica do pluralismo jurídico: a “teoria da polissistemia simultânea” de André-Jean Arnaud. *Quaestio Iuris*, 12(2), pp. 522-556.

Os marcos legais brasileiros para Internet, privacidade e proteção de dados pessoais representaram inegáveis avanços ao debate democrático, abrindo espaços para a reaproximação entre cidadãos e Estado (Beçak and Longhi, 2015, p. 138-139), no fenômeno mais amplamente conhecido como consensualização. Apesar disso, ainda são vistos com ceticismo no novíssimo debate sobre datacolonialismo (Faleiros Júnior and Borges, 2021b, pp. 25-31), devido aos incipientes mecanismos de sanção e à ainda incerta responsabilização por atos de controle exercidos nesse contexto e à dificuldade de imposição de limites à atuação de pequeno grupo de grandes corporações que atuam no norte global (Floridi, 2021).

É inegável que discutir criticamente a realidade sempre desafiadora da Infosfera é um passo importante em direção à abordagem ética desejada no desenvolvimento da IA, tendo em vista que, embora não se possa vislumbrar, antecipadamente, todas as consequências do implemento da técnica na sociedade da informação, ao menos é possível compreender alguns de seus problemas para que eventual mapeamento de riscos propicie a parametrização de estruturas de proteção que congreguem inovação e regulação em esperada harmonia, potencializada pelo pensamento descolonial como forma de rompimento do nefasto colonialismo digital de Inteligência Artificial imposto por pequenos grupos de empresa de tecnologia – as chamadas *Big Techs* (Rikap, 2021).¹² – que doutrinadores como Tim Wu defendem estar atingindo tamanho patamar de agigantamento que devem ser cindidas por leis antitruste.¹³

12 Rikap (2021) ilustra: “*The digital economy is highly asymmetric. By the time this is being written, its five leading corporations represent over 25% of the S&P 500. The combined market capitalization of Google, Apple, Facebook, Amazon and Microsoft (GAFAM) (5.587 trillion USD) is even above Japan’s 2019 GDP (5.01 trillion USD)*”.

13 Trata-se de uma aparente “nova” realidade, tamanha a empolgação gerada pelos algoritmos e por seu potencial, mas, segundo Tim Wu, as mesmas bases monopolistas que se observa agora já existiram no passado, e sempre foram marcadas por uma característica fulcral: o “tamanho” (*bigness*). Por isso o autor fala em uma nova “Era Dourada” (*a new Gilded Age*) e a elucida com a adjetivação trazida no título de sua obra, ao se referir à maldição (*curse*) que a grandeza de certas corporações propicia, pois é ela que produz os monopólios. Em proposta bastante polêmica, Wu ressalta que impor freios aos monopólios, no século XXI, deve ser uma prioridade tão grande quanto o foi em 1911, à época de Louis Brandeis – que o inspirou a cunhar o título da obra *The curse of bigness* – e da insatisfação geral com o monopólio do petróleo na expansão norte-americana para o Oeste. Entretanto, o desafio se tornou mais ardente que nunca, pois a tec-

Essa é a razão pela qual amplo debate vem sendo desenvolvido em torno do pensamento descolonial¹⁴ nos continentes do Hemisfério Sul.¹⁵ Quais são, pois, os limites que os direitos humanos podem impor para o enfrentamento de questões que envolvam a discriminação algorítmica na tutela de situações jurídicas existenciais na Internet? Como se pode frear a imposição oriunda de um novo colonialismo digital perpetrado de forma célere, mas quase imperceptível, tamanha sua sofisticação?! De fato, não é facultado ao Estado se furtar de oferecer proteção adequada a situações de discriminação, ainda que levadas a efeito na Internet, o que impõe o reconhecimento da premência do pensamento descolonial. Omissões regulatórias decorrentes do descompasso (anunciado) (Faleiros Júnior, 2020, pp. 64-66) entre a inovação tecnológica e a edição de leis escritas não podem representar *gaps* em ordenamentos que consagram diversas dimensões para a proteção dos direitos humanos (Maurer, 2007, p. 29), em suas inúmeras vertentes (Gottfried, 1999, p. 141).

São criados os “mercados ricos em dados” (*data-rich markets*), descritos por Viktor Mayer-Schönberger e Thomas Ramge como ambientes nos quais o usuário se torna espectador de suas próprias preferências, posto que seus dados são utilizados para mapear seus interesses e predizer suas deci-

nologia (e, em especial, a Internet) se introjeta no cotidiano social e produz exatamente o que deveria ser evitado: a aglutinação de “poder”. Diz: “*The Neo-Brandesian antitrust agenda is not an agenda for solving every economic challenge produced by the new Gilded Age. But structure matters, and these suggestions would help us return to an economic vision that prizes dynamism and possibility, and ultimately attunes economic structure to a democratic society.*” Wu, T. (2018) *The curse of bigness: antitrust in the new Gilded Age*. New York: Columbia Global Reports. p. 138.

- 14 Em sugestão de Catherine Walsh (2013, pp. 24-25), colhe-se a sugestão de utilização do termo “decolonial” ou “de-colonial”, sem a letra “s”, cuja supressão é uma opção adotada sinalizar distinção com o significado do prefixo “des” no castelhano, que poderia dar a entender um simples “[...] desarmar, desfazer ou reverter o colonial. [...]”. Com este jogo linguístico, procuro mostrar que não existe um estado nulo de colonialidade, mas sim posturas, posicionamentos, horizontes e projetos para resistir, transgredir, intervir, emergir, criar e influenciar”.
- 15 Ali, S. M. (2017) Decolonizing Information Narratives: Entangled Apocalypics, Algorithmic Racism and the Myths of History. *Proceedings of the Digitalisation for a Sustainable Society Summit*, 1(3), Article 50. doi: 10.3390/IS4SI-2017-03910. (Accessed: 10 January 2023); Birhane, A. (2020) Algorithmic colonization of Africa. *Scripted*, 17(2), pp. 389-409. doi: <http://doi.org/10.2966/scrip.170220.389>. (Accessed: 10 January 2023).

sões (Mayer-Schönberger and Ramge, 2018, p. 7).¹⁶ Tudo é funcionalizado a partir de uma nova commodity: a atenção.

Os “mercadores da atenção”, para citar a expressão de Tim Wu (2016, p. 5), são empresas que se valem de complexos algoritmos de Inteligência Artificial para disputar o interesse do público, obtendo cliques, gerando visualizações, adesões (seguidores, *likes* etc.)¹⁷ e, enfim, alterando a forma com que os indivíduos interagem e se relacionam. É uma vida on-line (ou “*onlife*” [on-line + *life*])¹⁸, como ilustra o acrônimo cunhado por Floridi (2015, pp. 1-6) potencializada pela predição estatística e pelos altíssimos lucros que esse novo modelo de mercado gera.¹⁹

Muito se fala em Inteligência Artificial, mas é fato que ainda não se atingiu o estado da técnica definido por Vernor Vinge como “singularidade tecnológica” (Vinge, 1993, pp. 11-22), em que verdadeira simbiose propiciará a indistinção entre humanos e máquinas no que concerne ao “jogo da imitação” proposto por Turing (1936)²⁰. O conceito de singulari-

16 Mayer-Schönberger and Ramge (2018, p. 7) comentam: “*The key difference between conventional markets and data-rich ones is the role of information flowing through them, and how it gets translated into decisions. In data-rich markets, we no longer have to condense our preferences into price and can abandon the oversimplification that was necessary because of communicative and cognitive limits*”.

17 Sobre o tema, conferir, por todos, Sumpter, D. (2018) *Outnumbered: From Facebook and Google to Fake News and Filter-bubbles – The Algorithms That Control Our Lives*. London: Bloomsbury Sigma; Vaidhyanathan, S. (2012) *The Googlization of Everything: (And Why We Should Worry)*. Berkeley: University of California Press.

18 Floridi (2015, pp. 1-6) explica: “*We decided to adopt the neologism “onlife” that I had coined in the past in order to refer to the new experience of a hyperconnected reality within which it is no longer sensible to ask whether one may be online or offline*”.

19 Com efeito: “*It is this larger narrative of transforming life through data that we are told is impossible to halt, because it is driven by a ‘technological momentum’ that is ‘inevitable.’ The notion that datafication is inevitable is, as we noted in chapter 1, a myth of data colonialism. But how to resist it? In this final part of the book, we try to wrest the narrative back from those who celebrate this transformation, and we question whether data colonialism’s emerging social and economic order is what human beings actually want*” (Couldry and Mejias, 2019, p. 188).

20 A ideia de uma máquina capaz de pensar leva a reflexões sobre os clássicos estudos de Alan Turing acerca do *Entscheidungsproblem* (dilema da tomada de decisão). Porém, o que Turing desconsiderou neste estudo foi justamente a possibilidade do ardil, da mentira e da dissimulação – comportamentos tipicamente humanos – e, ao delinear as premissas de seu famoso “Teste de Turing”, notou que potenciais erros de identificação poderiam surgir. Seria necessário avançar rumo à consolidação de um modelo mais complexo, por ele batizado de “Jogo da Imitação”, para que fosse possível aferir o potencial de uma máquina que pretenda, de forma convincente, “se passar por humana” (Turing, 1936). Ainda

dade, densamente explorado por Ray Kurzweil (1999, p. 213), indicaria um momento de emancipação; algoritmos se tornariam tão complexos que ultrapassariam os limites da predição estatística e do singelo suporte à tomada de decisões e, efetivamente, se tornariam máquinas capazes de “pensar” – sendo, por isso, artificialmente inteligentes – e com propensão à tomada de decisões “morais”.²¹

Como anota Dora Kaufman (2019, pp. 89-90),

[a]o longo de décadas, os modelos estatísticos tradicionais tentaram, por meio de várias metodologias, minimizar o grau de incerteza das previsões, enfrentando mais obstáculos do que resultados satisfatórios.

Basicamente, os algoritmos seguem evoluindo e se tornando tão complexos, em razão do avanço dos processos de *machine* e *deep learning*, que parecem propensos a superar os limites da simples predição estatística e do singelo suporte à tomada de decisões em função da heurística computacional para, efetivamente, se tornarem máquinas capazes de “pensar” – ao menos na leitura “futurista” que se faz sobre isso.

Nos dizeres de Jean-Gabriel Ganascia (2017, p. 13):

De acordo com Kurzweil, muito em breve estaremos baixando nossa consciência para máquinas, o que nos fornecerá uma forma de imortalidade. Isso resultará inevitavelmente da aceleração do progresso. Com efeito, segundo ele, a lei geral a que obedece toda forma de evolução, seja o desenvolvimento biológico, o aperfeiçoamento das civilizações ou o aperfeiçoamento das tecnologias, é de natureza intrinsecamente exponencial.²²

sobre o tema, conferir Doss, A. F. (2020) *Cyber Privacy: Who Has Your Data and Why You Should Care*. Dallas: BenBella Books, pp. 298-303.

21 Refletindo sobre os impactos desse fenômeno para a condição humana, conferir, por todos, Madsbjerg, C. (2017) *Sensemaking: the power of the humanities in the age of the algorithm*. New York: Hachette; Fry, H. (2018) *Hello world: how to be human in the age of the machine*. New York: Doubleday.

22 No original: “Selon Kurzweil, nous téléchargerons très bientôt notre conscience sur des machines, ce qui nous procurera une forme d’immortalité. Cela résultera fatalement de l’accélération des progrès. En effet, d’après lui, la loi générale à laquelle obéit toute forme d’évolution, qu’il s’agisse du développement biologique, du perfectionnement des civilisations ou de l’amélioration des technologies, est de nature intrinsèquement exponentielle” (Ganascia, 2017, p. 13).

Eduardo Tomasevicius Filho (2018, p. 142) enfatiza que “parece ser excessivamente artificial – ou mera ficção científica – imaginar um robô androide perfeitamente inteligente, andando pelas ruas sem qualquer controle.” Por óbvio, até que esse estágio de desenvolvimento seja atingido, o debate continuará a permear os limites do desenvolvimento tecnológico (Lloyd, 2011, p. 19) baseado na programação e nas condicionantes imponíveis/exigíveis dos desenvolvedores que, antecipada e conscientemente, definem o código-fonte que irá reger o funcionamento do algoritmo.²³ O debate permanece, por isso, no campo da predição, ao qual é pertinente a instigante perquirição dos riscos do desenvolvimento, de suas aplicações (cada vez mais consideradas em perspectiva nanotecnológica²⁴), além da imperiosa imposição de freios regulatórios.

Vale lembrar, citando Caitlin Mulholland, que, “para que uma pessoa seja obrigada a reparar um dano injusto, é fundamental que ela tenha a autonomia de atuação, isto é, tenha a capacidade de reconhecer a licitude ou ilicitude de sua conduta e, ao mesmo tempo, a habitualidade de identificar e prever a potencialidade danosa desta” (Mulholland, 2019, p. 332). Se a atribuição de personalidade jurídica a robôs não é solução

23 É exatamente por isso que tanto se discute a proposta de atribuição de personalidade eletrônica a robôs (operados por algoritmos de Inteligência Artificial). Na Europa, foi lançado o *Draft Report with recommendations on civil law rules and robotics* (2015/2103), do qual é possível extrair algumas preocupações quanto aos danos causados por máquinas. A partir desse documento, em 16 de fevereiro de 2017, o Parlamento Europeu aprovou uma Resolução (“Disposições de Direito Civil sobre Robótica”) que, expressamente e em caráter prospectivo, prevê em sua diretriz 59, “f”, a possibilidade de atribuição de personalidade eletrônica a robôs. O tema é complexo e instigou reflexões como as de Jacob Turner, que concebe a personalidade jurídica como mera ficção gerenciada por humanos, que podem atribuí-la a entes inumanos para finalidades pragmáticas. Turner, J. (2018) *Robot rules: regulating Artificial Intelligence*. Londres: Springer, p. 175. Uma síntese das principais preocupações relacionadas ao tema ainda pode ser obtida da leitura de Souza, E. N. de. (2020) Dilemas Atuais do Conceito Jurídico de Personalidade: Uma Crítica às Propostas de Subjetivação de Animais e de Mecanismos de Inteligência Artificial. *Civilistica.com*, 9(2), pp. 1-49, p. 40 e seg. E, finalmente, enaltecendo uma proposta inovadora e de cariz ‘intermediário’, extraída da doutrina alemã, tem-se o trabalho de Faleiros Júnior, J. L. de M. and Menke, F. (2020, August 6). “Teilrechtsfähigkeit”: uma proposta alemã para a responsabilização civil na IA. *Migalhas de Responsabilidade Civil* [online]. Available at: <https://s.migalhas.com.br/S/8AF9D>. (Accessed: 10 January 2023).

24 Sobre o tema, ver, por todos, Borges, G. S. and Martins, P. S. (2017). As nanotecnologias, os riscos incertos e a hipervulnerabilidade do nanoconsumidor. *Revista de Direito do Consumidor*, 113, 417-438.

viável, ou mesmo definitiva, cogita-se, noutro polo do espectro de possibilidades, do potencial dos seguros obrigatórios.

O debate, então passa a se concentrar na parametrização de deveres, especialmente no contexto delimitado pelos princípios da prevenção e da precaução, bem como de outros, como reversibilidade, segurança e responsabilidade (Barbosa, 2017, pp. 1501-1502). Entretanto, a responsabilidade civil não basta para solucionar todas as contingências decorrentes do implemento desses algoritmos em mercados complexos e datificados (Cave, 2020).

Como arrematam Eduardo Tomasevicius Filho e Angelo Viglianisi Ferraro (2020, p. 412), “chegou o momento de o problema da Inteligência Artificial ser seriamente abordado no nível político-regulatório”²⁵. E o arcabouço protetivo dos direitos humanos tem o condão de oferecer justamente isso: limites, a partir de parametrizações metanormativas, ao desenvolvimento algorítmico (Guerra Filho and Cantarini, 2021, p. 355).

3. A proteção aos direitos humanos e os limites à inovação desregulada

Nos dizeres de Mayer-Schönberger e Cukier (2014, p. 184), toda informação coletada em tempos de *Big Data* “não passa de um simulacro da realidade, como as sombras na parede da caverna de Platão.”²⁶ Quando se trabalha com fluxos massivos de dados, então, grandes preocupações passam a se introjetar na sociedade, não apenas quanto aos riscos de eventual uso discriminatório dos acervos de dados, mas até mesmo quanto ao surgimento de eventual/potencial dependência em relação a eles e às práticas de coleta massiva e de mineração de dados (*data mining*).²⁷

25 Tradução livre. No original: “Sono maturi i tempi perché il problema dell’Intelligenza artificiale venga affrontato seriamente a livello politico-normativo” (Tomasevicius Filho and Ferraro, 2020, p. 412).

26 Tradução livre. No original: “[...] It can only be a simulacrum of reality, like the shadows on the wall of Plato’s cave” (Mayer-Schönberger and Cukier, 2014, p. 184).

27 Gutwirth and Hildebrandt (2010, p. 33) afirmam: “Before embarking on the commonly known caveats regarding human rights like privacy, non-discrimination and due process, we like to stress the risks of an increased dependence on data mining technologies.

Os avanços da inovação e o ritmo acelerado do desenvolvimento de aplicações baseadas em algoritmos de Inteligência Artificial, de forma geral, também abrangem aspectos culturais, econômicos e políticos da vida na sociedade da informação, se introjetando em diversas estruturas centrais ao funcionamento social, despertando olhares para a perspectiva descolonial.²⁸ Assim, para suplantiar qualquer desdobramento capaz de acarretar regresso, especialmente quanto à proteção jurídica que se deve conferir aos direitos humanos²⁹, caminhos devem ser mapeados para conciliar inovação e regulação.³⁰ O pluralismo jurídico global surge, nesse aspecto, como vetor da função promocional dos direitos humanos³¹, abrindo espaços à tutela subjacente-valorativa da pessoa, mesmo em

Profiling is a powerful technique that renders visible what is invisible to the naked human eye. This, however, concerns patterns in databases that must not be mistaken for reality. By making visible what is aggregated in the data base, profiling also make invisible what cannot be translated into machine-readable data. In as far as the governance of people and things becomes dependent on these advanced profiling technologies, new risks will emerge in the shadow of the real time models and simulations these technologies make possible”.

- 28 Mohamed, Png and Isaac (2020, p. 7) anotam: “By recognizing the analogues of territorial and structural coloniality in the digital age, we propose the application of decolonial theory to digital technologies such as AI. Digital spaces—created by the Internet and the increasingly networked systems and devices we use—form digital territories that, like physical spaces, have the propensity to become sites of extraction and exploitation, and thus the sites of digital-territorial coloniality. Digital-structural coloniality also manifests, through to the coloniality of power, the coloniality of power can be observed in digital structures in the form of socio-cultural imaginations, knowledge systems and ways of developing and using technology which are based on systems, institutions, and values which persist from the past and remain unquestioned in the present. As such, emerging technologies like AI are directly subject to coloniality, giving decolonial critical theories a powerful analytical role”.
- 29 José de Oliveira Ascensão (2008, p. 278) aduz que os direitos do homem (direitos humanos, em sentido amplo), quando positivados em documentos internacionais de proteção e promoção da pessoa humana são considerados direitos humanos; quando positivados nas Cartas Constitucionais são considerados direitos fundamentais; e quando positivados na legislação civil são direitos de personalidade.
- 30 Couldry and Mejias (2019, p. 191) anotam: “To sum up, the problem and challenge of contemporary data practices is neither data nor simply the particular platforms that have emerged to exploit data”.
- 31 Conferir, por todos, Sarlet, I. W. (2010) *A Eficácia dos Direitos Fundamentais: Uma Teoria Geral dos Direitos Fundamentais na Perspectiva Constitucional*. 10th edn. Porto Alegre: Livraria do Advogado, p. 79; Hunt, L. (2009) *A invenção dos direitos humanos: uma história*. Translated by Rosaura Eichenberg. São Paulo: Cia. das Letras, pp. 113-145; Comparato, F. K. (2010) *A Afirmação Histórica dos Direitos Humanos*. 7th edn. São Paulo: Saraiva, pp. 91-92; Recaséns Siches, L. (2008) *Filosofia del derecho*. México: Porrúa, pp. 1-19.

um ambiente permeado pela disrupção tecnológica (Borges and Faleiros Júnior, 2021, p. 50).

O atingimento desse desiderato, na esteira do que defende Gustavo Zagrebelsky (1995, p. 15), somente ocorrerá se determinadas condicionantes estruturais se fizerem presentes, das quais a “*ductibilidad*” (maleabilidade) dos ordenamentos jurídicos constitucionais é a mais relevante, pois propicia o pacifismo e a integração democrática “através da rede de valores e procedimentos comunicativos que é, ademais, a única visão possível e não catastrófica da política em nosso tempo.”³² Nesse contexto, segundo Jorge Pereira da Silva:

O desiderato a se atingir é o de que o poder de intervenção estatal e a liberdade dos cidadãos se equilibrem de modo a garantir ao indivíduo tanta protecção quanto a necessária, mas também tanta liberdade pessoal quanto seja possível. Por isso, segundo a denominada concepção pessoal do bem jurídico, tem-se entendido que integram este conceito aquelas “realidades ou fins que são necessários para uma vida social livre e segura, que garantam os direitos humanos e fundamentais do indivíduo, assim como para o funcionamento do sistema estatal erigido para a consecução de tal objectivo. Não que, com esta referência, se pretenda induzir à importação acrítica para o direito constitucional dos resultados (nem sempre pacíficos) atingidos pela doutrina penalista sobre a teoria do bem jurídico – até porque a protecção penal é apenas uma modalidade, entre várias outras, de protecção de direitos fundamentais –, mas é importante reconhecer que a multifuncionalidade dos direitos fundamentais implica uma atenção redobrada ao conceito de bem jusfundamental e a sua colocação no centro do processo construtivo dos *conglomerados jurídicos* usualmente designados por direitos fundamentais (Silva, 2015, p. 354).

Embora não se possa deixar de considerar os impactos que as peculiaridades culturais acarretam para qualquer coletividade, a ponto de ser precipitada uma análise conjectural baseada na ideia de sociedade (ou “aldeia”³³) global, do ponto de vista dos direitos humanos, posições identitárias e individuais impõem a ponderação (Herrera Flores, 2009, pp. 97-98),

32 Tradução livre. No original: “*a través de la red de valores y procedimientos comunicativos que es además la única visión no catastrófica de la política posible en nuestro tiempo*” (Zagrebelsky, 1995, p. 15).

33 Cf. McLuhan and Fiore (1971).

notadamente para que sejam fixadas firmemente as bases do entrelaçamento entre direito público e direito privado no sentido da superação do colonialismo digital (Ali, 2014, pp. 28-35).

A Inteligência Artificial afeta exatamente esse contexto, pois distorce os tradicionais *standards* de proteção historicamente construídos (Greene, Park and Colaresi, 2019, pp. 223-230). Altera-se, essencialmente, a própria acepção que se tem do que é considerado “humano” (Coleman, 2019). Tudo se reduz à matemática, à heurística, à predição e ao mapeamento dos interesses como condicionantes funcionais de estruturas de mercado, pois não se atingiu a já descrita singularidade tecnológica. As máquinas não são capazes de agir ou tomar decisões imbuídas de concepções morais.

Em um universo no qual a predição algorítmica está presente de forma tão marcante, nichos de aglutinação de poder desenvolvem ambientes menos seguros à proteção dos direitos humanos. Nesse sentido, Shoshana Zuboff (2019b, p. 28) fala na instrumentação e instrumentalização do comportamento para fins de modificação, previsão, monetização e controle ao propor o termo “instrumentarismo” (*“instrumentarianism”*)³⁴, que simboliza o epítome do que a própria autora designa como capitalismo de vigilância. O poder instrumental, em seus dizeres, realiza a expropriação da experiência humana como um imperativo econômico, processando decisivamente a redistribuição dos direitos humanos elementares dos indivíduos para o capital.

Nesse campo, o colonialismo digital se torna intrinsecamente dependente de dados, e ganha espaço na medida em que se torna extensão de um processo global de extração que começou sob o colonialismo e continuou através do capitalismo industrial, mas que, atualmente, em vez de recursos naturais e trabalho, se apropria da vida humana por meio de sua

34 Zuboff (2019b, p. 28) explica: “*As to the new species of power, I have suggested that it is best understood as instrumentarianism, defined as the instrumentation and instrumentalization of behavior for the purposes of modification, prediction, monetization and control. In this formulation, “instrumentation” refers to the ubiquitous, sensate, computational, actuating global architecture that renders, monitors, computes, and modifies human behavior. Surveillance capitalism is the puppet master that imposes its will through the vast capabilities of this connected puppet to produce instrumentarian power, replacing the engineering of souls with the engineering of behavior.*”

conversão em dados e do estímulo incessante à produção de maior volume de dados coletáveis.

Estruturas de monitoramento e vigilância³⁵ permitem tais extrações e degradam a vida, tornando-a datificada (Harari, 2018, p. 88) e convertendo-a em produto direto para fins de produção capitalista (Couldry and Mejias, 2019, p. 9). Assim, uma das consequências fundamentais desse fenômeno é o rastreamento contínuo da vida humana, cada vez mais rastreada (e rastreável)!

Para frear esse indesejado paradigma é preciso, naturalmente, reconfigurar estruturas protetivas condizentes com o novo momento do desenvolvimento técnico-informacional (Walsh, 2018). Os direitos humanos devem inspirar marcos regulatórios, propostas legislativas e, essencialmente, todo o acervo normativo que se pretenda instituir.

No Brasil, o que guiou a promulgação da Lei Geral de Proteção de Dados Pessoais foi justamente esse “núcleo duro” de parâmetros extraídos de uma compreensão mais ampla do direito à privacidade (Trstenjak, 2016, p. 48).³⁶, previstos no texto legal como fundamentos, em seu artigo 2º, e que atuam como vetores axiológicos para os direitos (e deveres) descritos noutras passagens da lei e, também, para a atuação posterior do Estado, no exercício de seu poder regulatório. Destaca-se a redação do inciso VII do artigo 2º da LGPD: “A disciplina da proteção de dados pessoais tem como fundamentos: (...) VII – os direitos humanos, o livre desenvolvimento da personalidade, a dignidade e o exercício da cidadania pelas pessoas naturais.”

A definição de categorias merecedoras de maior proteção, como a dos dados pessoais adjetivados como “sensíveis” (art. 5º, II, da LGPD) é evidência sólida dessa preocupação do legislador. Outra evidência disso é estruturação de revisões das decisões automatizadas (art. 20, da LGPD), que devem ser realizadas por agentes humanos. Se a proteção insuficiente

35 Conferir: Marx, G. T. (1984) *Fragmentation and cohesion in American society*. Washington, DC: Trend Analysis Program.

36 Trstenjak (2016, p. 48): “*The right to privacy functions as a ‘window’ to the constitutional rights, which protects the private and personal sphere of an individual, e.g., dignity, free development of personality, religious freedom, privacy and data protection.*”

não pode ser admitida, sob pena de flagrante violação ao citado fundamento da lei, deve-se estruturar mecanismos de controle que atuem como freios aos desideratos que afastem os humanos de sua essência (Moyn, 2018, p. 220).³⁷

Cass Sunstein (2014, p. 137) elucida o problema:

Quando as pessoas usam heurísticas simples, ou atalhos mentais, geralmente é porque funcionam bem, no sentido de que nos permitem tomar boas decisões. Mas mesmo que as heurísticas geralmente funcionem bem, elas podem levar a grandes erros. Quando fazemos avaliações imprecisas de probabilidades, pode muito bem ser porque heurísticas simples estão nos levando ao erro.³⁸

Que fique claro: o recrudescimento valorativo dos direitos humanos não implica considerar um resgate antropocêntrico, egoístico ou que coloque o homem (individualmente considerado) novamente no centro do sistema jurídico – como foi no Estado Liberal –, ou seja, não é o homem econômico (*homo economicus*) a figura que se pretende ver inserida no vértice constitucional, ainda que este também seja merecedor de proteção pontual, a nível fundamental (Faleiros Júnior and Borges, 2021a, pp. 297-300).

Almeja-se, sim, a maior proteção ao “homem existencial”, concebido a partir da proteção de experiências individuais que tenham uma projeção útil para o próprio titular e para a coletividade. É nesse contexto que se colhe o maior valor da delimitação de fundamentos normativos nos dispositivos introdutórios da norma. Postulados instituídos com tal cariz atuam para além da lei especificamente considerada e inspiram o ordenamento como um todo.

37 Moyn (2018, p. 2020) assevera: “Human rights will return to their defensible importance only as soon as humanity saves itself from its low ambitions. If it does, for the sake of local and global welfare, sufficiency and equality can again become powerful companions, both in our moral lives and in our political enterprises”.

38 Tradução livre. No original: “When people use simple heuristics, or mental shortcuts, it is generally because they work well, in the sense that they enable us to make good decisions. But even if heuristics usually work well, they can lead to big errors. When we make inaccurate assessments of probabilities, it may well be because simple heuristics are leading us astray” (Sunstein, 2014, p. 137).

4. Considerações finais

O mencionado descompasso entre inovação e regulação cria um ambiente não regulado no qual os agentes de dados, que buscam a todo custo conquistar mercados, podem levar à exploração excessiva desses dados por algoritmos avançados, o que é uma tendência perigosa porque pode resultar em violações aos direitos humanos. É necessário diferenciar entre perigos e riscos para que possam ser propostas regulamentações baseadas na identificação de padrões extralegais para preencher as “áreas cinzentas” de exploração que desafiam o Direito a encontrar soluções para o complexo problema criado por modelos algorítmicos que consideram os interesses da maioria.

Os direitos humanos não podem ser violados e sua proteção deve ser um objetivo geral para qualquer agente que trate dados pessoais. A lei que protege os direitos humanos não se limita a um único artigo, mas é uma parte importante de todo o sistema jurídico e define limites para o desenvolvimento descontrolado. Mesmo que os dados possam ser coletados e tratados para fins legais sem o consentimento, é importante que as violações sejam punidas com rigor. Nesse sentido, pelo que se observou no curso da pesquisa, a hipótese lançada se confirma, denotando a necessidade de que direitos amplamente reconhecidos com tal cariz sejam preservados em detrimento de estruturas algorítmicas que os possa violar, ainda que sob o manto da inovação ou da livre iniciativa.

O papel da conceituação atribuída aos dados pessoais sensíveis pela legislação de regência propicia manancial exemplificativo de direitos que ostentam essa característica e que devem receber maior proteção do ordenamento, não se esgotando, portanto, no rol do artigo 5º, II, da LGPD. Ademais, a própria instituição da proteção aos direitos humanos como fundamento da referida lei (art. 2º, VII) denota a preocupação do legislador com a construção de uma estrutura de postulados que ultrapassa a lei individualmente considerada e inspira todo o ordenamento, conectando fontes e definindo freios ao desenvolvimento desregulado.

Nesse sentido, ainda que possam ser coletados e até mesmo tratados para finalidades lícitas – inclusive com dispensa de consentimento, nas hipóteses apresentadas pela lei –, é imperioso que eventuais violações sejam punidas com mais severidade. Igualmente, é fundamental que sua preservação seja vista como um objetivo geral de todo agente de tratamento de dados. A proteção aos direitos humanos, como dito, atua exatamente nesse sentido, oferecendo um incremento nas condições de identificação de espaços de maior propensão a falhas e danos.

Impõe-se a leitura crítica quanto à possibilidade de rompimento ou desconstrução do malfadado colonialismo digital que já é exercido por empresas tecnológicas do Norte Global. Assim, TICs que estão em franca expansão, como os algoritmos de Inteligência Artificial, devem ser contemplados em vista de sua propensão à colonialidade do poder, o que demanda estudos sobre sua descolonização a fim de que, quando desenvolvido, permita atenuar as injustiças sociais decorrentes do capitalismo de vigilância.

Referências

- Ali, S. M. (2017) ‘Decolonizing Information Narratives: Entangled Apocalypitics, Algorithmic Racism and the Myths of History’. *Proceedings of the Digitalisation for a Sustainable Society Summit*, 1(3), Article 50. doi: 10.3390/IS-4SI-2017-03910. (Accessed: 10 January 2023).
- Ali, S. M. (2014) ‘Towards a decolonial computing’ in Autónoma University of Lisbon (ed.) *Ambiguous Technologies: Philosophical Issues, Practical Solutions, Human Nature, International Society of Ethics and Information Technology*, pp. 28-35.
- Ascensão, J. O. (2008) ‘A dignidade da pessoa e o fundamento dos direitos humanos’. *Revista da Faculdade de Direito da Universidade de São Paulo*, 103, pp. 277-299.
- Barbosa, M. M. (2017) ‘Inteligência artificial, e-persons e direito: desafios e perspectivas’. *Revista Jurídica Luso-Brasileira*, 3(6), pp. 1475-1503.

- Beçak, R. and Longhi, J. V. R. (2015) 'Abertura e colaboração como fundamentos do Marco Civil da Internet: a atuação do Poder Público na construção do governo eletrônico brasileiro e a governança da Internet' in De Lucca, N., Simão Filho, A. and Lima C. R. P. de (eds.) *Direito & Internet III: Marco Civil da Internet (Lei nº 12.965/2014) (Vol. I)*. São Paulo: Quartier Latin.
- Birhane, A. (2020) 'Algorithmic colonization of Africa'. *Scripted*, 17(2), pp. 389-409 [online]. Available at: <http://doi.org/10.2966/scrip.170220.389> (Accessed: 10 January 2023).
- Borges, G. S. and Faleiros Júnior, J. L. M. (2021) 'Decolonial thinking in Brazil: perspectives for overcoming digital colonialism through the protection of human rights'. *Studia Prawnicze: Rozprawy i Materiały*, 2(29), pp. 43-53.
- Borges, G. S. and Martins, P. S. (2017) 'As nanotecnologias, os riscos incertos e a hipervulnerabilidade do nanoconsumidor'. *Revista de Direito do Consumidor*, 113, pp. 417-438.
- Bragato, F. F. (2014) 'Para além do discurso eurocêntrico dos direitos humanos: contribuições da descolonialidade'. *Revista Novos Estudos Jurídicos*, 19(1), pp. 201-230.
- Cave, S. (2020) The Problem with Intelligence: Its Value-Laden History and the Future of AI. *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics and Society*, pp. 29-35.
- Cave, S. and Dihal, K. (2020) 'The Whiteness of AI'. *Philosophy & Technology*, 33, pp. 685-703 [online]. Available at: <http://doi.org/10.1007/s13347-020-00415-6>. (Accessed: 10 January 2023).
- Coleman, F. (2019) *A Human Algorithm: How Artificial Intelligence is Redefining Who We Are*. Berkeley: Counterpoint.
- Comparato, F. K. (2010) *A Afirmação Histórica dos Direitos Humanos*. 7th edn. São Paulo: Saraiva.
- Couldry, N. and Mejias, U. A. (2019) *The Costs of Connection: How Data is Colonizing Human Life and Appropriating it for Capitalism*. Stanford: Stanford University Press.

- Domingos, P. (2015) *The Master Algorithm: How the Quest for the Ultimate Learning Machine will Remake Our World*. New York: Basic Books.
- Doss, A. F. (2020) *Cyber Privacy: Who Has Your Data and Why You Should Care*. Dallas: BenBella Books.
- Eubanks, V. (2018) *Automating inequality: How high-tech tools profile, police and punish the poor*. New York: St. Martin's Press.
- Faleiros Júnior, J. L. de M. (2020) *Administração Pública digital: Proposições para o aperfeiçoamento do Regime Jurídico Administrativo na sociedade da informação*. Indaiatuba: Foco.
- Faleiros Júnior, J. L. de M. and Borges, G. S. (2021a) 'Ética algorítmica e direitos humanos: reflexões sobre os limites do "profiling" no capitalismo de vigilância' in Cantarini, P., Guerra Filho, W. S. and Séllos Knoerr, V. C. (eds.) *Direito e inteligência artificial: fundamentos, Vol. 2: Inteligência artificial e tutela de direitos* (pp. 293-330). Rio de Janeiro: Lumen Juris.
- Faleiros Júnior, J. L. de M. and Borges, G. S. (2021b) 'Inteligência artificial e datacolonialismo' in Beçak, R. and Bernardo de O. C. G. (coords.), *Inteligência artificial e democracia: desafios no Brasil do século XXI*. Belo Horizonte: Arraes.
- Faleiros Júnior, J. L. de M. and Menke, F. (2020) "'Teilrechtsfähigkeit": uma proposta alemã para a responsabilização civil na IA'. *Migalhas de Responsabilidade Civil*, 6 August [online]. Available at: <https://s.migalhas.com.br/S/8AF9D> (Accessed: 10 January 2023).
- Floridi, L. (2010) *Information: A very short introduction*. Oxford: Oxford University Press.
- Floridi, L. (2015) Introduction in L. Floridi (ed.) *The onlife manifesto: being human in a hyperconnected era*. Cham: Springer, pp. 1-6.
- Floridi, L. (2021) Las redes sociales no se pueden regular solas. *El País*, 26 January [online]. Available at: <https://elpais.com/opinion/2021-01-25/las-redes-sociales-no-se-pueden-regular-solas.html> (Accessed: Jan. 10, 2023).

- Fry, H. (2018) *Hello world: how to be human in the age of the machine*. New York: Doubleday.
- Gajewski, M. and Krawiec, P. (2016) 'Identification and access to objects and services in the IoT environment' in Mavroumoustakis C., Mastorakis G. and Mongay Batalla, J. (eds.) *Internet of Things (IoT) in 5G Mobile Technologies*. Cham: Springer, pp. 275-298.
- Ganascia, J.-G. (2017) *Le mythe de la singularité: faut-il craindre l'intelligence artificielle?* Paris: Editions du Seuil.
- Gottfried, P. E. (1999) *After liberalism: mass democracy in the managerial State*. Princeton: Princeton University Press.
- Greene, K. T., Park, B. and Colaresi, M. (2019) 'Machine learning human rights and wrongs: how the successes and failures of supervised learning algorithms can inform the debate about information effects'. *Political Analysis*, 27(2), pp. 223-230.
- Greengard, S. (2015) *The Internet of Things*. Cambridge: The MIT Press.
- Guerra Filho, W. S. and Cantarini, P. (2021) 'Por uma teoria inclusiva do direito digital: tutela de direitos fundamentais em face da inteligência artificial' in Cantarini, P., Guerra Filho, W. S. and Séllos-Knoerr, V. C. (eds.), *Direito e inteligência artificial: fundamentos, Vol. 2: Inteligência artificial e tutela de direitos*. Rio de Janeiro: Lumen Juris, pp. 349-366.
- Gutwirth, S. and Hildebrandt, M. (2010) 'Some caveats on profiling' in Gutwirth, S., Pouillet, Y. and Hert, P. de (eds.), *Data protection in a profiled world*. Cham: Springer, pp. 31-42.
- Harari, Y. N. (2018) *21 lições para o século 21*. Translated by Paulo Geiger. São Paulo: Cia. das Letras.
- Hayles, N. K. (2012) 'How we think: transforming power and digital technologies' in Berry, D. M. (ed.) *Understanding digital humanities*. London: Palgrave Macmillan, pp. 42-66.
- Herrera Flores, J. (2009) *Teoria crítica dos direitos humanos: os direitos humanos como produtos culturais*. Translated by Luciana Caplan. Rio de Janeiro: Lumen Juris.

- Hui, Y. (2019). *Recursivity and contingency*. London: Rowman & Littlefield. E-book.
- Hunt, L. (2009) *A invenção dos direitos humanos: uma história*. Translated by Rosaura Eichenberg. São Paulo: Cia. das Letras.
- Kaufman, D. (2019) *A inteligência artificial irá suplantará a inteligência humana?* Barueri: Estação das Letras e Cores.
- Kearns, M. and Roth, A. (2020) *The ethical algorithm: the science of socially aware algorithm design*. Oxford: Oxford University Press.
- Kurzweil, R. (1999) *The age of spiritual machines: when computers exceed human intelligence*. New York: Viking.
- Lanier, J. (2013) *Who owns the future?* New York: Simon & Schuster.
- Lloyd, I. J. (2011) *Information technology law*. 6th edn. Oxford: Oxford University Press.
- Madsbjerg, C. (2017) *Sensemaking: the power of the humanities in the age of the algorithm*. New York: Hachette.
- Marx, G. T. (1984) *Fragmentation and cohesion in American society*. Washington, DC: Trend Analysis Program.
- Maurer, H. (2007) *Contributos para o direito do estado*. Translated by Luís Afonso Heck. Porto Alegre: Livraria do Advogado.
- Mayer-Schönberger, V. and Cukier, K. (2014) *Big data: a revolution that will transform how we live, work, and think*. New York: Houghton Mifflin Harcourt.
- Mayer-Schönberger, V. and Ramge, T. (2018) *Reinventing capitalism in the age of Big Data*. New York: Basic Books.
- McLuhan, H. M. and Fiore, Q. (1971) *Guerra e paz na aldeia global*. Translated by Ivan Pedro de Martins. Rio de Janeiro: Record.
- Mignolo, W. D. (2017) 'Colonialidade: o lado mais escuro da modernidade'. Translated by Marco Oliveira. *Revista Brasileira de Ciências Sociais*, 32(94), pp. 1-18.

- Mohamed, S., Png, M.-T. and Isaac, W. (2020) 'Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence', *Philosophy & Technology*, 33, pp. 659-684 [online]. Available at: <http://doi.org/10.1007/s13347-020-00405-8> (Accessed: 10 January 2023).
- Moyn, S. (2018) *Not enough: human rights in an unequal world*. Cambridge: Harvard University Press.
- Mulholland, C. (2019) 'Responsabilidade civil e processos decisórios autônomos em sistemas de Inteligência Artificial (IA): autonomia, imputabilidade e responsabilidade' in Frazão A. and Mulholland C. (eds.) *Inteligência Artificial e Direito: ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, pp. 325-348.
- Muntandas, B. (2021) 'Algoritmos en la vida cotidiana: apps, gadgets y dependencia tecnológica' in Barbosa, M., Braga Netto, F., Silva, M. and Faleiros Júnior, J. L. de M. (eds.) *Direito digital e inteligência artificial: diálogos entre Brasil e Europa*. Indaiatuba: Foco, pp. 641-658.
- O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown.
- Peppet, S. R. (2014) 'Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security and Consent', *Texas Law Review*, 93(1), pp. 85-176.
- Quijano, A. (2005) *Colonialidade do Poder. A Colonialidade do Saber: Eurocentrismo e Ciências Sociais. Perspectivas Latino-americanas*. Buenos Aires: CLACSO.
- Recaséns Siches, L. (2008) *Filosofia del derecho*. México: Porrúa.
- Rikap, C. (2021) *Capitalism, Power and Innovation: Intellectual Monopoly Capitalism Uncovered*. London: Routledge.
- Rouland, N. (2003) *Nos Confins do Direito: Antropologia Jurídica da Modernidade*. Translated by M. E. de Almeida Prado Galvão. São Paulo: Martins Fontes.
- Sarlet, I. W. (2010) *A Eficácia dos Direitos Fundamentais: Uma Teoria Geral dos Direitos Fundamentais na Perspectiva Constitucional*. 10th ed. Porto Alegre: Livraria do Advogado.

- Silva, J. P. da (2015) *Deveres do Estado de Proteção de Direitos Fundamentais: Fundamentação e Estrutura das Relações Jusfundamentais Triangulares*. 3rd ed. Lisboa: Universidade Católica Editora.
- Silveira, S. A. da (2020) 'Sistemas Algorítmicos, Subordinação e Colonialismo de Dados' in Sabariego, J., Amaral, A. J. do and Salles, E. B. C. (eds.), *Algoritarmos*. São Paulo: Tirant Lo Blanch, pp. 160-172.
- Souza, E. N. de (2020) 'Dilemas Atuais do Conceito Jurídico de Personalidade: Uma Crítica às Propostas de Subjetivação de Animais e de Mecanismos de Inteligência Artificial' *Civilistica.com*, 9(2), pp. 1-49.
- Sumpter, D. (2018) *Outnumbered: From Facebook and Google to Fake News and Filter-bubbles – The Algorithms That Control Our Lives*. London: Bloomsbury Sigma.
- Sunstein, C. R. (2014) *Valuing Life: Humanizing the Regulatory State*. Chicago: The University of Chicago Press.
- Thatcher, J., O'Sullivan, D. and Mahmoudi, D. (2016) 'Data Colonialism through Accumulation by Dispossession: New Metaphors for Daily Data'. *Environment and Planning D: Society and Space*, Preprint, pp. 1-17. doi: 10.1177/0263775816633195. (Accessed: Jan. 10, 2023).
- Tomasevicius Filho, E. (2018) 'Inteligência Artificial e Direitos da Personalidade: Uma Contradição em Termos?' *Revista da Faculdade de Direito da Universidade de São Paulo*, 113(1), pp. 133-149.
- Tomasevicius Filho, E. and Ferraro, A. V. (2020) 'Le nuove sfide dell'umanità e del diritto nell'era dell'Intelligenza Artificiale'. *Revista Direitos Culturais*, 15(37), pp. 401-413.
- Trstenjak, V. (2016) 'General report: The influence of human rights and basic rights in private law' in Trstenjak, V. and Weingerl, P. (eds.) *The influence of human rights in private law*. Cham: Springer.
- Turing, A.M. (1936) 'On computable numbers, with an application to the Entscheidungsproblem'. *Proceedings of the London Mathematical Society*, 42(1), pp. 230-265.

- Turner, J. (2018) *Robot rules: regulating Artificial Intelligence*. Londres: Springer.
- Vaidhyanathan, S. (2012) *The Googlization of Everything: (And Why We Should Worry)*. Berkeley: University of California Press.
- Vanderlinden, J. (1989) 'Return to legal pluralism: twenty years later'. *Journal of Legal Pluralism*, 28, p. 149-157.
- Villas Bôas Filho, O. (2019) 'Uma abordagem sociológica do pluralismo jurídico: a "teoria da polissistemia simultânea" de André-Jean Arnaud', *Quaestio Iuris*, 12(2), pp. 522-556.
- Vinge, V. (1993) 'The coming technological singularity: How to survive in the post-human era' in Vinge, V. *Interdisciplinary Science and Engineering in the Era of Cyberspace*, p. 11-22. NASA John H. Glenn Research Center at Lewis Field, Cleveland [online]. Available at: <https://ntrs.nasa.gov/search.jsp?R=19940022856> (Accessed: Jan. 10, 2023).
- Walsh, C. (2018) 'Insurgency and Decolonial Prospect, Praxis, and Project' in Mignolo, W. D. and Walsh C. (eds.) *On decoloniality: Concepts, analytics, praxis*. Durham: Duke University Press, pp. 33-56.
- Walsh, C. (2013) *Pedagogías decoloniales: prácticas insurgentes de resistir, (re)existir y (re)vivir*, vol. 1. Quito, Ecuador: Ediciones Abya-Yala.
- Wu, T. (2016) *The attention merchants: the epic scramble to get inside our heads*. New York: Vintage.
- Wu, T. (2018) *The curse of bigness: antitrust in the new Gilded Age*. New York: Columbia Global Reports.
- Zagrebelsky, G. (1995) *El derecho dúctil. Ley, derechos y justicia*. Translated by Marina Gascón. Madrid: Trotta.
- Zeifert, A. P. B. and Agnoletto, V. (2019) 'O pensamento descolonial e a teoria crítica dos direitos humanos: saberes e dignidade nas sociedades latino-americanas'. *Revista Húmus*, 9(26), pp. 197-218.
- Zuboff, S. (2019a) *The age of surveillance capitalism: the fight for a human future at the new frontier of power*. New York: Public Affairs.

Zuboff, S. (2019b) “‘We make them dance’: surveillance capitalism, the rise of instrumentarian power, and the threat to human rights’ in Jørgensen, R. F. (ed.), *Human rights in the age of platforms*. Cambridge: The MIT Press, p. 3-52.

INTELIGÊNCIA ARTIFICIAL E DISCRIMINAÇÃO ALGORÍTMICA: MARCOS REGULATÓRIOS E PARÂMETROS ÉTICOS

Haide Maria Hupffer¹

Gustavo da Silva Santanna²

DOI: 10.29327/5312711.1-3

Resumo: Decisões algorítmicas impulsionadas por técnicas de *Machine Learning* (ML) podem discriminar grupos legalmente protegidos, acirrar preconceitos, perpetuar desigualdades, violar os direitos humanos ou criar novas formas de injustiça. Ferramentas automatizadas usadas para seleção de candidatos à emprego, concessão de empréstimo, liberação de seguro, educação, contratação de plano de saúde e outras formas de decisões automatizadas habilitadas por inteligência Artificial têm se mostrado com vieses discriminatórios e preconceituosos, o que passa a exigir políticas mitigadoras e novas leis. A maioria dos algoritmos de Inteligência Artificial ainda operam como caixas pretas que fornecem pouca ou nenhuma explicação sobre as decisões tomadas. A proteção legal positiva na Constituição Federal de não discriminação é desafiada quando a Inteligência Artificial, e não os humanos, discrimina. O

- 1 Pós-Doutora e Doutora em Direito pela Unisinos. Pesquisadora no Programa de Pós-Graduação em Qualidade Ambiental e no Curso de Direito da Universidade Feevale. Líder do Grupo de Pesquisa CNPq/Feevale Direito e Desenvolvimento. Este capítulo é parte do resultado da pesquisa produzida no âmbito do seguinte projeto de investigação científica: “Inteligência Artificial e Sociedade de Algoritmos: regulação, riscos discriminatórios, governança e responsabilidades”, financiado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul – Fapergs – Processo número 61271.674.22274.26112021, Edital Fapergs 07/2021 – Programa Pesquisador Gaúcho – Pq. ORCID: <https://orcid.org/0000-0002-4965-9258>. Lattes: <http://lattes.cnpq.br/4950629941533824> E-mail: haide@feevale.br.
- 2 Doutor e Mestre em Direito pela Unisinos. Professor da Universidade Federal do Rio Grande do Sul – UFRGS – e da Atitus Educação. Professor das Especializações em Direito do Estado da Universidade Federal do Rio Grande do Sul – UFRGS, em Direito Digital, Gestão da Inovação e Propriedade Intelectual e Direito Administrativo da Pontifícia Universidade Católica de Minas Gerais – PUC/MINAS. Procurador do Município de Alvorada/RS. Advogado. ORCID: <https://orcid.org/0000-0003-1190-3355>. Lattes: <http://lattes.cnpq.br/3394849903197236>. E-mail: gssantanna@hotmail.com.

artigo tem como objetivo principal identificar algumas externalidades negativas e os impactos que algoritmos programados como vieses discriminatórios e preconceituosos podem produzir na sociedade e aos direitos fundamentais, bem como busca-se analisar os principais movimentos regulatórios para lidar com os desafios éticos e legais que envolvem a questão da discriminação algorítmica, responsabilidade pela programação em todas as fases do ciclo de vida, transparência, segurança, explicabilidade, impactos nos direitos fundamentais e nos ecossistemas, reafirmando o direito à dignidade e à não discriminação e os valores da igualdade e da justiça. A pesquisa é exploratória e descritiva, com apoio no método dedutivo e na análise documental e revisão da literatura.

Palavras-chave: Discriminação Algorítmica. Vieses Discriminatórios. Explicabilidade.

Sumário: 1. Introdução. 2. Inteligência Artificial: algumas aplicações. 3. As múltiplas faces da discriminação invisível impulsionada por algoritmos em sistemas de inteligência artificial. 4. O desafio de regulamentar a Inteligência Artificial para garantir a promoção da igualdade e a não discriminação. 5. Conclusão. Referências.

1. Introdução

A Inteligência artificial (IA) está revolucionando todas as áreas do conhecimento com mudanças profundas na forma como o ser humano está vivendo e trabalhando. No entanto, como acontece com qualquer tecnologia, existem limitações e desafios a serem considerados para que opere com segurança, responsabilidade, respeito ao ser humano e que se utilize a tecnologia desenvolvida em decisões para o bem e não para o mal. Todas as áreas estão sendo profundamente impactadas e a cada nova inovação, velhas questões continuam sem respostas, dentre as quais citam-se: quem controla um sistema de IA se sua capacidade de aprendizagem leva a decisões que não fo-

ram as intenções projetadas na sua configuração inicial? Se a IA tomar uma decisão discriminatória, de quem é a responsabilidade pelas consequências legais, morais e sociais? Se for delegado às máquinas autonomia plena para tomada de decisão e livre arbítrio, quem garante que sistemas de IA ao adquirirem sua própria consciência não gerem seus próprios vieses?

Dotar a IA com capacidade de tomada de decisão autônoma e capacidade de aprendizado passa a exigir estruturas éticas, legais e de governança. Sistemas de IA precisam de muitos dados e parâmetros para serem treinados com precisão, contudo deve-se ter presente que os dados são alimentados e escolhidos por humano, e possíveis vieses discriminatórios podem ser transferidos gerando decisões racistas e sexistas.

A partir do exposto, o presente estudo objetiva identificar algumas externalidades negativas e os reflexos que algoritmos enviesados podem produzir na sociedade e aos direitos fundamentais, bem como examinar os principais movimentos regulatórios para lidar com as espinhosas questões éticas, sociais e políticas sobre os riscos da discriminação algorítmica para que a IA não perpetue o racismo estrutural enraizado nas decisões humanas.

A pesquisa é de natureza qualitativa e exploratória, adota o método dedutivo e utiliza como procedimento técnico a pesquisa bibliográfica e documental. A pesquisa documental é apoiada nas diretivas da União Europeia, Unesco, OCDE e de alguns países que avançaram na elaboração de diretivas no que concerne a responder aos desafios específicos colocados pela discriminação algorítmica, com aportes éticos e normativos que estabelecem salvaguardas mínimas para a proteção contra a discriminação algorítmica, respeito aos direitos humanos, justiça, transparência, segurança, prestação de contas e respeito aos valores democráticos em todo o ciclo de vida de sistemas de IA.

2. Inteligência Artificial: algumas aplicações

Com a finalidade de prolongar seu ciclo de vida, os algoritmos aprendem a interpretar interativamente as tarefas

a ele submetidas (classificação, regressão, árvore de decisão, modelos estatísticos, processos matemáticos, agrupamento, métodos bayesianos para inferência estatísticas e ANNs). Os algoritmos e os sistemas de IA se adaptam e utilizam a aprendizagem de máquina (*machine learning*) para alcançar objetivos, encontrar insights ocultos e padrões sem serem explicitamente programados para tal. Dito de outro modo, a máquina aprende (*machine learning*) com bases em dados de cálculos/respostas anteriores e melhoram com a experiência, podendo auxiliar na construção de decisões confiáveis e repetíveis. A *machine learning* é uma aplicação da IA que permite o aprendizado e o aperfeiçoamento de sistemas conduzidos por grandes volumes de dados, sendo já utilizados em diferentes áreas, como para auxiliar em prognósticos de doenças, na pontuação de crédito, na detecção de fraude, no reconhecimento facial e reconhecimento de voz, no processamento da linguagem natural (PNL) (Janiesch and Zschoenck, 2021).

Com base no problema apresentado e nos dados disponíveis, as etapas necessárias para treinar e implantar *machine learning* são diferentes, como observam Lloyd, Mohseni e Rebentrost (2013). Os autores apontam três tipos de *machine learning*: i] supervisionada (a máquina infere uma função a partir de um conjunto de dados que foram treinados para classificar e realizar previsões – o modelo vai se ajustando de forma interativa para avaliar diferentes características até alcançar o resultado desejado); ii] não supervisionada (quando não há supervisão do sistema, mas apenas os dados de entrada são fornecidos – o modelo de aprendizagem não supervisionado ajuda a encontrar tipos de padrões e estruturas ocultas em dados não rotulados, concentrando-se, em particular, no problema de aprendizagem de máquinas em larga escala – *big data* – em que tanto o treinamento como o número de recursos são grandes, o que possibilita uma assertiva maior na segmentação de clientes e uma comunicação mais efetiva com o público-alvo); iii] aprendizagem por reforço (aplicada com grande sucesso em ambientes fechados como em jogos e para sistemas multiagentes) (Lloyd, Mohseni and Rebentrost, 2013).

Startups estão explorando o uso de *machine learning* na indústria farmacêutica para projetar novos medicamentos e com promessas de serem muito mais seguros e eficazes. Os primeiros medicamentos para câncer projetados com o auxílio da Inteligência Artificial (IA) estão agora na fase de rigorosos testes clínicos com voluntários para observar se funcionam e se o tratamento é seguro para, em sequência, solicitarem aos órgãos reguladores a liberação para uso generalizado. Como exemplo, Heaven (2023) cita que desde o ano de 2021, a empresa Exscientia (Reino Unido) com apoio da Universidade de Medicina de Viena (Áustria) e outras empresas farmacêuticas estão testando uma nova tecnologia de *matchmaking*, que desenvolve medicamentos precisos para tratamento de câncer que é individualizado para cada paciente, levando em conta diferenças biológicas sutis entre as pessoas. Coletam amostras do tecido do paciente, dividem a amostra para incluir em uma amostra células normais e na outra amostra de células cancerígenas, dividindo-a novamente em mais cem partes e, em sequência, expõem cada uma das partes a vários coquetéis de drogas. Depois de expostos, utilizam a automação robótica e *machine learning* para observar o que acontece com a exposição de cada um dos coquetéis (Heaven, 2023).

Com a tecnologia de *matchmaking*, a IA possibilita realizar uma busca exaustiva pelo medicamento certo, sem precisar submeter o paciente ao tratamento convencional que exige vários meses de quimioterapia. Nos experimentos realizados, os pesquisadores perceberam que alguns medicamentos não mataram as células cancerígenas do paciente observado e em outros experimentos prejudicaram células saudáveis. Com o processo de *matchmaking*, o paciente recebeu o medicamento correto contra o seu câncer e dois anos após o tratamento o câncer desapareceu e estava em remissão completa. A indústria farmacêutica está apostando muito, pois além da IA propiciar descobertas de medicamentos de forma muito mais rápida, também pode se tornar mais barata e acessível a todos os pacientes. A aprendizagem de máquina reduz o trabalho meticuloso para o desenvolvimento de novos medicamentos em laboratórios (Heaven, 2023).

A técnica de aprendizado profundo (*deep learning*) possibilita a modelos computacionais compostos por várias camadas de processamento apreenderem com a descoberta de estruturas complexas de grandes conjuntos de dados usando o algoritmo de retropropagação. Este algoritmo é usado para “indicar como uma máquina deve alterar seus parâmetros internos para calcular a representação em cada camada a partir da representação na camada anterior”. O *deep learning* propicia avanços no processamento de falas, imagens, imagens, vídeo e áudio, enquanto as redes recorrentes iluminaram os dados sequenciais, como texto e fala” (LeCun, Bengio and Hinton, 2015). A área da medicina utiliza abordagens baseadas em *deep learning*, como *deep neural networks*, que faz parte de uma família mais ampla de técnicas de aprendizado de máquina baseadas em redes neurais artificiais, alcançando resultados impressionantes no processamento de imagens de inúmeras doenças. As abordagens de *deep learning* são inspiradas na capacidade do cérebro humano de abstrair representações de alto nível com estímulos sensoriais de baixo nível (Hossain et al, 2023).

Os exemplos da medicina mostram que estão sendo realizados esforços substanciais para o enriquecimento de aplicações de IA com imagens médicas usando *deep learning* para diagnosticar erros em sistemas de diagnósticos de doenças, prever sintomas precoces de doenças, reconhecer e extrair padrões, melhorar a precisão de diagnósticos para uma variedade de doenças com utilização de algoritmos supervisionados ou não supervisionados. *Machine* e *deep learning* em estudos de dados médicos multidimensionais são utilizados para endossar o processo de decisão usando imagens médicas que apresentam, como toda nova tecnologia, vantagens e desvantagens (Latif et al, 2019).

A combinação entre IA, *machine learning*, *analytics* e *big data* está revolucionando todas as áreas pela possibilidade de auxiliar gestores e tomadores de decisão pela necessidade de mais previsibilidade sobre os resultados de cada ação tomada. A IA consegue cruzar dados, “identificar relações e gerar *insights* em um nível que seria praticamente impossível para

o cérebro humano”. Setores tradicionais, como o agronegócio, estão usando a IA para detectar doenças e pragas, realizar uma estimativa de safra, identificar fenômenos que influenciam nas atividades de plantio, automatizar a contagem de frutas, máquinas agrícolas autônomas e inteligentes, previsão meteorológica personalizada para cada fazenda, distinguir entre frutos verdes de maduros e agricultura de precisão (MIT Technology Review, 2022).

Drones são desenvolvidos para serem usados massivamente em conflitos armados, com capacidade para encontrar seus próprios alvos e atacar infraestruturas sensíveis, o que os tornam apenas mais uma questão de programação de algoritmos, com consequências catastróficas se mal utilizados. A integração da IA generativa nas tecnologias de consumo podem ampliar os efeitos dos atuais sistemas de IA, tanto de forma positiva como negativa, podendo criar textos e imagens

A sensação do momento é o ChatGPT que foi lançado em novembro de 2022 pela OpenAI trouxe a Inteligência Artificial para o debate cotidiano. O que atrai os olhares é sua habilidade de resolver problemas complexos e com mais precisão do que os anteriores, atraindo em apenas um mês a marca de mais de 100 milhões de usuários. É um algoritmo que processa a linguagem humana e gera conteúdos utilizando modelos que foram treinados com GPTs (*Generative Pre-Trained Models*). Outra característica que explica o seu sucesso é o de ser uma interface conversacional mais acessível aos usuários leigos e a enorme base de dados de conhecimento que possui. Os algoritmos são treinados para poder aprender, usando o melhor dos conceitos de aprendizagem de máquina e redes neurais profundas. Por isso, em linguagem mais simples é conhecido como sendo uma mistura de um robô conversacional com algoritmos do tipo generativo aqueles que conseguem gerar conteúdo (Del Rey, 2023).

Neurotecnologias, por serem mais invasivas, geram preocupações éticas e de governança, pois podem adentrar na privacidade e na liberdade cognitiva. São inúmeros os avanços na área de neurotecnologia, como as que permitem o monitoramento da atividade cerebral para análises e pesquisas,

decodificares de humor, tecnologias capazes de ler a mente, bem como identificar culpados na área de segurança. Por isso, a necessidade de criar mecanismos, normativas e fronteiras que estabeleçam limites (MIT Technology Review, 2023a). Outra inovação, conhecida como interface subvocal, possibilita incorporar em óculos inteligentes com uso de IA uma interface para reconhecer comandos não vocalizados apenas com os movimentos dos lábios e da boca, que podem ser executados em conjunto com um celular. No futuro essa tecnologia poderá devolver a vós para os pacientes. Atualmente, ela já é usada para comunicações com outras pessoas em espaços onde falar poderia ser inconveniente ou inapropriado, a exemplo de bibliotecas ou locais barulhentos (Inovação Tecnológica, 2023).

O uso de algoritmos e da Inteligência Artificial para tomada de decisões está presente em uma ampla gama de setores, desde planejamento de tráfego, concessão de plano de saúde, diagnóstico de doenças, filtragem de *spam*, reconhecimento de fala e reconhecimento facial, assertividade para pagamento de empréstimos, se o candidato à vaga será um bom funcionário, gerenciamento de logística, previsão de alvos prováveis para intervenção policial, prevenir crimes ou resolver crimes passados fazendo previsões estatísticas (Borgesius, 2020). Contudo, a tomada de decisão por algoritmos pode colocar em risco os direitos humanos, como o direito à não discriminação, como pontua Borgesius (2020): “embora a tomada de decisão algorítmica possa parecer racional, neutra e imparcial, ela também pode levar à discriminação injusta e ilegal”.

Uma decisão algorítmica é definida por Borgesius (2020) como o processo pelo qual o algoritmo produz uma saída (*output*). Dependendo da configuração do sistema de IA, o algoritmo pode decidir de forma totalmente automatizada em processos que antes eram de responsabilidade exclusiva de humanos, como também é utilizado para preparar decisões, ou seja, quando os humanos tomam decisões com base nos algoritmos. Em geral, é difícil de mensurar os limites entre a tomada de decisão humana e a automatizada. Os problemas de processo de tomada de decisão são múltiplos e complexos

e os riscos para indivíduos, grupos e sociedade são semelhantes para decisões totalmente ou parcialmente automatizadas (Council Of Europe, 2018).

Operada por meio de algoritmos, a IA assume uma importante função de realizar análises preditivas, baseadas na análise de dados de informações alimentadas por *big data*, mas muito se discute se esse processo pode interferir na autonomia e na privacidade, se será tendencioso e discriminatório, se poderá ser contaminado com erros e ser excessivamente invasivo. Uma preocupação geral é a falta de transparência desses processos o que passa a exigir medidas precaucionais para mitigar o risco (Zarsky, 2013, p. 1506).

3. As múltiplas faces da discriminação invisível impulsionada por algoritmos em sistemas de inteligência artificial

O uso de algoritmos de Inteligência Artificial, *Machine* e *deep learning* tornaram-se comuns no dia a dia das pessoas, além dos benefícios amplamente divulgados, podem gerar vieses contra minorias, mulheres e outras classes protegidas. Ao longo dos anos inúmeros países aprovaram leis para proteger os consumidores contra a discriminação na concessão de crédito, habitação e emprego, onde reguladores e agências têm a tarefa de fazer cumprir essas leis. Uma das questões mais prementes na adoção de algoritmos é a dificuldade de garantir justiça e transparência na sua utilização. Sistemas de recrutamento e seleção podem estar configurados para dar notas mais baixas para mulheres com pele mais escura, que se graduaram em determinada faculdade e que moram em determinado bairro ou que participaram de movimentos feministas; da mesma forma pela localização geográfica ou geografia da mídia social. Minorias relacionadas à origem racial ou imigrantes podem ser propensas a receberem menos ofertas de marketing relacionadas a moradias (Schmidt and Stephens, 2019, pp. 130-131).

Há uma crescente preocupação e debates sobre os riscos e benefícios que essas inovações tecnológicas represen-

tam, por serem considerados pouco transparentes e reprodutoras de desigualdades e racismo. Se mal programadas e usadas são capazes de perpetuar injustiças tendenciosas e resultar em decisões imprecisas, discriminatórias, bem como violar os direitos humanos fundamentais. Nos Estados Unidos, a Casa Branca, o *Federal Reserve Board*, o *Consumer Financial Protection Bureau* (CFPB), o *Government Accountability Office* e a *Federal Trade Commission* divulgam relatórios e frequentemente solicitam informações sobre pautas relacionadas à justiça algorítmica quanto ao uso da Inteligência Artificial, *Machine* e *deep learning* nos mercados de emprego, crédito e habitação, e no uso de dados de consumidores, funcionários e mercado (Schmidt and Stephens, 2019, pp. 130-131).

Quando o sistema é desenvolvido com viés discriminatório ou quando aprendeu com decisões humanas discriminatórias e tendenciosas, as decisões algorítmicas resultam em mais discriminação e em grave ofensa aos direitos humanos. Como exemplo, registra-se a utilização de decisões de algoritmos para policiamento com base em estatísticas criminais alimentadas com viés discriminatório. Se o sistema policial prestar mais atenção em um bairro onde se concentram muitos imigrantes, ele passará a registrar mais crimes daquele bairro do que em outro lugar. Esse registro não ocorre porque realmente há um volume maior criminalidade naquele local, mas pelo fato de deslocar mais policiais e, conseqüentemente, fazer o registro de todas as abordagens. O policiamento preditivo tem o potencial de causar um ciclo de feedback extremamente perigoso, pois os algoritmos podem representar decisões humanas discriminatórias, o que resulta em novas decisões preconceituosas que ferem os direitos humanos. (Borgesius, 2020).

Schmidt e Stephens (2019, p. 138) levantam o seguinte questionamento: o que significa um algoritmo ser justo? Os autores reportam que os algoritmos de *machine learning* treinados e os dados nos quais eles se baseiam para tomar decisões podem não ser isentos de viés, visto que são capazes de refletir os resultados de um processo de decisão tendencioso (humano ou não). Da mesma forma, um modelo bem intencionado

é capaz de criar um algoritmo que eterniza o viés existente achado nos dados utilizados para treinar o algoritmo. Igualmente, Schmidt e Stephens (2019, p. 138) chamam a atenção para o fato de que “algoritmos imprecisos ou mal construídos podem também captar correlações espúrias em dados que inadvertidamente levam a resultados relativamente piores para diferentes grupos demográficos”. Por isso, o alerta dos autores sobre a importância de a sociedade reconhecer que é possível a ocorrência de vieses algorítmicos, que dados recém disponibilizados, mesmo que estejam associados à confiança, não estão livres de conterem vieses discriminatórios. Por essa razão, torna-se premente um grande esforço para desenvolver novos métodos e implementar abordagens que possam garantir que algoritmos utilizados em processos de tomada de decisão sejam tão justos quanto possível (Schmidt and Stephens, 2019, p. 144).

No âmbito da e-Administração, o uso da Inteligência Artificial pode incorporar as mazelas de uma Administração Pública Patrimonialista, ou seja, pessoalizada, e, consequentemente, discriminatória. Se sabe que muitas das dificuldades enfrentadas hoje em termos de gestão pública advêm das vicissitudes de um modelo de gestão que deveria já ter sido ultrapassado. Diante disso, não podem essas distorções ser incorporadas aos sistemas robotizados. Um exemplo desse problema pode ser extraído do COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), um sistema de IA utilizado em estágios sequenciais da justiça criminal nos Estados Unidos em vários Estados (Wisconsin, Arizona, Delaware, Colorado, Louisiana, Kentucky, Washington, Virgínia), que inclui prejulgamento e correções comunitárias, liberdade condicional, avaliação de risco de reincidência de réus, auxiliando magistrados na dosimetria de penas e na concessão de liberdade condicional. O COMPAS considera uma série de variáveis sobre o réu (*input*) e é alimentado pelo próprio réu quando é efetuada uma prisão, com informações sobre histórico de violência, histórico criminal, abuso de substâncias, local de moradia, histórico escolar, profissão, ambiente social, lazer, se algum familiar já foi preso ou está preso, envolvimen-

to com pessoas que pertencem a organizações criminosas, problemas financeiros. Esses dados inseridos pelo réu são utilizados para contabilizar a porcentagem de reincidência. As respostas são inseridas no software COMPAS para gerar várias pontuações, inclusive análise preditiva de “Risco de reincidência” e “Risco de reincidência violenta” (Larson et al, 2016).

Larson et al (2016) analisaram mais de 10.000 réus criminais do Condado de Broward, Flórida, e compararam as taxas de reincidência previstas pela ferramenta COMPAS com as taxas reais de reincidência dos réus em um período de dois anos. Na análise realizada, os pesquisadores desvelaram que os réus negros, por este sistema, eram muito mais propensos do que os réus brancos a serem julgados incorretamente como estando em maior risco de reincidência, enquanto os réus brancos tinham um escore mais baixo de risco. Na análise realizada estava correta a previsão de reincidência do infrator em 61% das vezes, mas estava correta apenas em suas previsões de reincidência violenta 20%. Em relação a prever quem poderia reincidir em crime, o algoritmo previu “corretamente a reincidência para réus negros e brancos aproximadamente na mesma taxa (59% para réus brancos e 63% para réus negros), mas cometeu erros de maneiras muito diferentes. Ele classifica erroneamente os réus brancos e negros de maneira diferente quando examinados ao longo de um período de acompanhamento de dois anos”. Nas análises os pesquisadores também descobriram que “réus negros que não reincidiram em um período de dois anos tinham quase duas vezes mais chances de serem erroneamente classificados como de maior risco em comparação com seus colegas brancos (45% contra 23%)”. Restou claro aos pesquisadores que réus negros eram erroneamente classificados como possíveis reincidentes em crimes violentes (Larson et al, 2016). É um exemplo emblemático de um sistema de IA desenvolvido com algoritmos tendenciosos com pontuações de risco que eram racialmente injustas, o que gera discriminação algorítmica e reproduz preconceitos enraizados na sociedade.

Na área de recrutamento e seleção empresas estão preocupadas com possíveis vies inconsciente projetados na

construção de algoritmos em sistemas de IA com preconceitos que estão enraizados no cotidiano. A *startup* brasileira Jobecam fundada em 2019 está desenvolvendo uma plataforma que possibilita que o candidato ao emprego e o recrutador se concentrem apenas nas habilidades e competências exigidas para a vaga. À vista disso, a *startup* alterou sua plataforma “que permite filtrar currículos e realizar entrevistas, pré-gravadas ou ao vivo, de forma totalmente anônima”. O objetivo não é esconder o candidato ao emprego, mas sim dar ênfase às reais habilidades e competências do candidato. Outra forma utilizada para minimizar a discriminação é não inserir o ano da formação universitária ou de curso profissionalizante para “evitar os cálculos e driblar o etarismo, disponibiliza apenas o período de cada um”. Outra opção é possibilitar para a contratante ocultar os nomes da universidade na qual o candidato se graduou e o nome das empresas em que já trabalhou, para minimizar discriminação (Cardial, 2023).

Críticas de que os algoritmos não são neutros e que podem representar novas formas de preconceitos ou exacerbar as já existentes dependerá em grande medida das escolhas que são feitas na concepção de sistemas de IA. Muitas das escolhas são moldadas pelo ambiente ético e legal em que a empresa opera. Contudo, não é tão simples capturar os dados ameaçados por algoritmos tendenciosos que violam a legislação (Kim, 2017, p. 865). Empregadores tendenciosos escondem sua intenção discriminatória em algoritmos, podendo até intencionalmente desejarem que determinado resultado discriminatório ocorra e, assim, usam modelos que mascaram decisões discriminatórias por trás da fachada neutra de análise de dados. No geral, a vítima tem dificuldade de provar ter sido discriminada. Por isso, é preciso ficar atento em relação as escolhas na codificação da informação pelos programados de sistemas de IA, “erros nos dados, confiança em amostras não representativas, ou a seleção de variáveis para exclusão ou inclusão pode produzir um modelo impreciso de maneira sistemática”, para evitar que esses erros sistemáticos reduzam ainda mais as oportunidades de grupos desfavorecidos (Kim, 2017, p. 884).

Uma outra forma de gerar discriminação algorítmica é quando um erro no registro de um indivíduo (redes sociais, dados coletados de sites públicos) pode sugerir que ele seja inadimplente em um empréstimo ou que possa ter um antecedente criminal, quando na verdade essa informação que está na rede não é verdadeira. “Informações imprecisas não aumentam inerentemente preocupações de igualdade, pois os erros podem ser distribuídos aleatoriamente, infectando os registros de membros de grupos privilegiados, bem como de grupos protegidos” (Kim, 2017, pp. 885-886).

Se o algoritmo fizer uma previsão e tomar uma decisão com base em erros de dados da pessoa, pode privar injustamente o indivíduo de oportunidades de emprego, acesso à crédito, acesso a plano de saúde e seguro, por exemplo.

Essa “discriminação invisível” ou “discriminação por associação” pode pôr em xeque a ideia de os algoritmos serem possuidores de “neutralidade” (Navarro, 2017, pp. 48-51). Isso porque as condutas estabelecidas pelos robôs que utilizam inteligência artificial o são com base em padrões de comportamentos gerados pelos próprios seres humanos e que a podem levar à discriminação com base na raça, etnia, idade, localização geográfica, renda, etc. O Estado-Administração, a contrário senso, de base desses dados, deve, sim, direcionar suas políticas públicas a fim de otimizar seus serviços e custos, de maneira a transformar a “discriminação” em “igualdade social” (Navarro, 2017, pp. 48-51).

Na Alemanha, essa aceitação, que envolve diretamente elementos da personalidade, denomina-se: direito à “autodeterminação informativa” ou “autopresentação” que consiste no direito de o próprio indivíduo decidir acerca da divulgação e utilização de seus dados pessoais, possibilitando, inclusive, que o indivíduo “se insurja contra as falsas, não autorizadas, degradantes ou deturpadas representações de sua pessoa, bem como o protege das observações secretas e indesejadas de sua personalidade” (Menke, 2015, p. 210). Diferencia-se da autodeterminação que trata do direito de o indivíduo determinar/definir sua identidade, desde a origem biológica até a orientação sexual. Também se distingue da “autopreservação”

que consiste na garantia de o indivíduo recolher-se para si, de ficar só, abrangendo o sigilo a diários pessoais, boletins médicos e materiais biológicos (Menke, 2015, p. 210). Essa preocupação se deve, principalmente, em razão da utilização que pode ser dada aos dados dos indivíduos, uma vez que, estes podem ser manipulados por instituições (públicas ou privadas) sem que o indivíduo saiba disso (Menke, 2015, pp. 210-2011). É exatamente no combate à manipulação dos dados e à discriminação que pode ser gerada em razão das informações, que o respeito à privacidade se impõe.

Stefano Rodotà (2008, p. 35), inclusive, chega a sugerir que, se fosse possível redefinir uma classificação acerca da privacidade, o grau máximo de opacidade seria dado àquelas que pudessem gerar práticas discriminatórias, sendo de grau “máximo de transparência aquelas que, referindo-se à esfera econômica dos sujeitos, concorrem para embasar decisões de relevância coletiva.” Logo, a privacidade não possui caráter absoluto (Batalla, 2010, p. 767) e sua limitação deve estar condicionada a determinadas circunstâncias, como a exigência legal de seu abrandamento, para evitar eventuais “intromissões ilegítimas” (Limberger, 2007, pp. 127-130). Da mesma maneira os dados coletados deveriam ser mínimos, “*no sólo para preservar el derecho a la protección de datos, sino también para darles mayor efectividad*” (Batalla, 2010, p. 735).

Algoritmos são considerados os elementos-chave que sustentam serviços e infraestruturas cruciais na sociedade da informação. Ao serem projetados para proporcionar acesso amplo e igualitário à sociedade, cada vez mais a humanidade depende de algoritmos para tomar decisões importantes. Contudo, algoritmos não são eticamente neutros e podem acirrar ou perpetuar diferentes riscos de discriminação, como já registrado, e ameaçar direitos de grupos legalmente protegidos. São questões complexas que precisam ser enfrentadas pelo Direito e que exigem um grande esforço para desbloquear a “caixa preta” algorítmica para entender como esses sistemas tomam decisões (Xenidis and Senden, 2020, pp. 1-2). Programadores que desenvolvem sistemas podem também não saber como ele se comportará quando utilizado na prática e

quando for alimentado com determinados dados. De igual forma, segredos de Estado ou segredos comerciais podem dificultar a obtenção de informações sobre sistemas algorítmicos (Borgesius, 2020).

O foco principal dos desenvolvedores é garantir que os algoritmos executem as tarefas para os quais foram projetados. Compreender o tipo de “pensamento que orienta os desenvolvedores é essencial para entender o surgimento de vieses em algoritmos e na tomada de decisões algorítmicas”, como pontuam Tsamados et al (2022). Os autores observam que o pensamento que predomina no campo do desenvolvimento de algoritmos é o “formalismo algorítmico”, ou seja, os desenvolvedores aderem as regras e formas prescritas e se esforçam para serem neutros, mas esquecem que essa abordagem ignora a complexidade da sociedade e do mundo real, o que amplia o risco de consolidar as condições sociais existentes, como desigualdades estruturais que prejudicam etnias. Vieses algorítmicos podem ocorrer em qualquer estágio do processo, ou seja, desde o design, desenvolvimento, implantação e tomada de decisão. Tsamados et al (2022) alertam que podem ocorrer ciclos viciosos quando algoritmos realizam avaliações equivocadas de um determinado grupo no com base na mera correlação. Resultados injustos e com vieses ampliam a discriminação.

4. O desafio de regulamentar a Inteligência Artificial para garantir a promoção da igualdade e a não discriminação

Há consenso sobre a necessidade de avançar em busca da justiça algorítmica, em especial, para mitigar os riscos de discriminação direta (tratamento desigual) e indireta (impacto desigual) frente as decisões algorítmicas. Contudo, não há consenso sobre como definir, medir e estabelecer padrões de justiça algorítmica. Uma forma de mitigar os riscos é investir em pesquisas para esclarecer a natureza dos riscos éticos com orientação sólida e transparente para a governança do design dos algoritmos e do uso da tecnologia. (Tsamados et al, 2022).

Não é a tecnologia específica (computação em nuvem, celulares, computadores, plataformas online e assim por diante) que causa problemas éticos e discriminatórios, mas sim o que o hardware faz com software e com os dados em todo o ciclo de vida, desde o desenvolvimento de algoritmos, coleta de dados, compartilhamento, armazenamento, análise e uso dos dados. Logo, a macroética está se voltando para as diferentes dimensões morais da utilização de cada vez mais dados (pessoais sensíveis), questões de privacidade, anonimato, transparência, confiança, responsabilidade e com a crescente dependência de algoritmos para analisar os dados, a fim de moldar escolhas e tomar decisões. Outra questão presente nas discussões é a redução gradual do envolvimento humano nas decisões de algoritmos ou mesmo na supervisão de muitos processos automatizados, o que passa a exigir que sejam colocadas questões prementes de equidade, responsabilidade, respeito pelos direitos humanos e não discriminação (Floridi and Taddeo, 2016).

Floridi e Taddeo (2016) pontuam que os desafios éticos colocados pela ciência dos dados estão delineados em três eixos de pesquisa: a ética dos dados, a ética dos algoritmos e a ética das práticas. Para os autores, a ética dos dados diz respeito a coleta e análise de grandes conjuntos de dados (desde o uso de *big data* até pesquisas biomédicas, das áreas de ciências sociais, publicidade, filantropia de dados, criação de perfis), a possível reidentificação de indivíduos com a utilização de tecnologias de mineração de dados e com a reutilização de grandes conjuntos de dados, o que pode resultar em riscos para a chamada “privacidade de grupo” e sérios problemas éticos de discriminação de grupos (por idade, sexismo, preconceito, etnia) até violência contra grupos. No que concerne a ética dos algoritmos, especialmente no caso de ML (*machine learning*), os desafios éticos incluem “responsabilidade moral, responsabilidade do designer e dos cientistas de dados em relação as consequências imprevistas e indesejadas (por exemplo, discriminação ou promoção de conteúdo antissocial), bem como oportunidades perdidas”. Por sua vez, a ética das práticas está centrada no consentimento do titular de dados, na privacidade

do usuário e no uso secundário dos dados. Nessa linha, a ética das práticas preocupa-se com a “responsabilidade de pessoas e organizações encarregadas de processos, estratégias e políticas de dados, incluindo cientistas de dados, com o objetivo de definir uma estrutura ética para moldar códigos profissionais sobre inovação, desenvolvimento e uso responsáveis”. As práticas éticas devem garantir que os algoritmos e sistemas de IA possam promover “tanto o progresso da ciência de dados quanto a proteção dos direitos de indivíduos e grupos” (Floridi and Taddeo, 2016).

Um dos principais desafios na atualidade apontados por Morley et al (2020) relaciona-se ao que é nominado vantagem dupla de *machine learning* éticas, ou seja, como avançar para que as oportunidades sejam capitalizadas se ainda não estão claros os danos que alimentam todo o ciclo de vida dos algoritmos. Como construções sociotécnicas poderosas, os algoritmos de *machine learning* levantam inúmeras preocupações em relação aos desenvolvedores e organizações que os projetam quanto sobre seu código que pode ser a maior promessa como a maior ameaça à humanidade. Os autores argumentam que sem transparência e pela complexidade e dificuldade de interpretar os códigos será difícil controlar, monitorar e corrigir sistemas algorítmicos. Contudo, há um apelo coletivo de cientistas sociais, filósofos, eticistas, formuladores de políticas, tecnólogos e sociedade civil para que sejam desenvolvidos mecanismos de governança apropriados para a sociedade. Dito de outro modo, além de capitalizar as oportunidades é necessário ter certeza de que os valores éticos, princípios e os direitos humanos estão sendo respeitados e que a tomada de decisão seja justa e eticamente responsável (Morley et al, 2020).

A busca pela justiça algorítmica, responsabilidade e transparência, são temáticas cada vez mais presentes nas discussões sobre as implicações éticas dos algoritmos. A cada dia surgem novos problemas éticos e formas de abordá-los. Falar sobre viés algorítmico tornou-se um tema central para lidar com essa nova realidade digital e “suas consequências para o Estado Democrático de Direito, à ordem jurídica estabelecida e

aos princípios gerais do direito”. A União Europeia avançou no que concerne a responder aos desafios específicos colocados pela discriminação algorítmica, com aportes normativos que estabelecem salvaguardas mínimas para a proteção contra a discriminação nos Estados-Membros. O direito de não ser discriminado é um princípio geral e um direito fundamental no direito da União Europeia (Xenidis and Senden, 2020, pp. 1-2).

O art. 22 do *General Data Protection Regulation* (GDPR) proíbe certas decisões automatizadas incluindo a definição de perfis ao dispor 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”. Contudo, a proibição não se aplica para os seguintes casos: i] quando a decisão “for necessária para a celebração ou a execução de um contrato entre o titular dos dados e um responsável pelo tratamento”; ii] quando o titular dos dados deu consentimento à decisão automatizada; iii] quando for “autorizado por lei da União Europeia ou do Estado-Membro a que o responsável pelo tratamento estiver sujeito, e na qual estejam igualmente previstas medidas adequadas para salvaguardar os direitos e liberdades e os legítimos interesses do titular dos dados” (União Europeia, 2016).

Em situações em que o controlador puder confiar no consentimento informado do titular ou na exceção contratual uma regra diferente é acionada (Borgesius, 2020). Essa regra está prevista também no art. 22 da GDPR que deixa claro que o responsável pelo tratamento de dados tem o dever de aplicar “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” (União Europeia, 2016).

Como exemplo, Borgesius (2020) cita que um banco pode possibilitar que uma decisão automatizada seja reconsiderada ao negar automaticamente um empréstimo por meio de seu site. Nessa situação, o cliente pode ligar para o banco

solicitando que um funcionário reconsidere a decisão. Em relação a transparência para decisões automatizadas, a GDPR nos artigos 13, inciso 2 (f) e 14, inciso 2 (f) dispõe que o responsável pelo tratamento deve informar ao titular dos dados sobre a “existência de tomadas de decisão automatizadas, incluindo a definição de perfis” e neste caso o responsável deve dar informações claras, compreensíveis e “significativas sobre a lógica envolvida, bem como o significado e as consequências previstas de tal processamento para o titular dos dados” (Borgesius, 2020).

Diretivas éticas e documentos foram desenvolvidos por Estados, empresas, grupos de pesquisadores, universidades e organizações nos últimos anos com recomendações para adoção de princípios éticos para a IA. Registra-se que o primeiro padrão intergovernamental sobre IA foi publicado pela *Organisation for Economic Cooperation and Development* (OECD), que anunciou no dia 22 de maio de 2019 que seus trinta e seis países membros, juntamente com outros seis países parceiros (Argentina, Brasil, Colômbia, Costa Rica, Peru e Romênia), concordaram em estabelecer padrões éticos para a IA. O documento recomendatório contempla diretrizes e princípios da OECD (2019) que visam promover a inovação para que a IA seja confiável, assim como garantir que os algoritmos sejam projetados para serem justos, seguros, robustos e que respeitem os direitos humanos e os valores democráticos em todas as etapas do ciclo de vida. A recomendação se concentra em torno de cinco princípios: i] Crescimento inclusivo, desenvolvimento sustentável e bem-estar; ii] valores centrados no ser humano e na justiça; iii] transparência e explicabilidade; iv] robustez e segurança; v] prestação de contas (OECD, 2019).

Também no ano de 2019, Cingapura lançou a primeira edição do “*Model AI Governance Framework (Model Framework)*” que fornece orientação detalhada para empresas do setor privado com abordagens éticas e de governança importantes para implantar soluções de IA e, assim, promover a compreensão e confiança do público. Como princípios orientadores também elegem que as decisões tomadas pela IA devem ser explicáveis, transparentes, justas e centradas no ser humano. O fra-

network desenvolvido está alicerçado em cinco práticas que as empresas devem observar: i] criar estruturas de governança interna com funções e responsabilidades claras para monitorar e gerenciar riscos e treinar as equipes; ii] determinar o nível de envolvimento humano na tomada de decisão com grau apropriado para minimizar o risco de danos aos indivíduos; iii] atentar para o gerenciamento de operações para que sejam minimizados vieses nos dados e nos algoritmos, para atender esse requisito é importante medidas de redução de risco como a explicabilidade, robustez e possibilitar ajustes regulares; iv] interação e comunicação com uma rede ampla de partes interessadas, ou seja, as políticas de IA devem ser divulgadas para os usuários em linguagem compreensível, como também deve ser permitido que os usuários forneçam feedback; v] robustez e reprodutibilidade. Registra-se que no ano de 2018 foi criado o Conselho Consultivo sobre o Uso Ético de IA e Dados, cujos membros são nomeados pelo Ministro das Comunicações e da Informação. Em 4 de dezembro de 2020, Cingapura lançou o primeiro guia para auxiliar “organizações e funcionários a entenderem como as funções de trabalho existentes podem ser redesenhadas para aproveitar o potencial da IA, de modo que o valor de seu trabalho seja aumentado” e os riscos minimizados (Singapore, 2022).

Outro documento importante foi publicado pela União Europeia no ano de 2020 com o título *The ethics of artificial intelligence: Issues and initiatives*. É um extenso documento que sistematiza os principais dilemas éticos e as questões morais associadas à implantação da IA, levantando questões sobre robôs assumindo papéis de humanos nos relacionamentos, manipulação intencional do mercado, problemas legais relacionadas às decisões tomadas por IAs, responsabilidade por incitar atividades criminosas, assédio, tortura, crimes sexuais, fraude, responsabilidade por acidentes com veículos autônomos, direitos humanos e bem-estar, danos emocionais, risco existencial, dano ambiental e sustentabilidade, dano social e justiça social, dentre outros temas. O documento também apresenta diversas iniciativas nacionais e internacionais sobre ética para a IA nos seguintes temas: direitos

humanos, bem-estar das pessoas, impactos da IA na experiência emocional e na manipulação emocional e intencional dos usuários, cultural, perpetuação de injustiças e possibilidade de ampliar a discriminação contra grupos específicos da sociedade, recomendações de estruturas de governança, necessidade de prestação de contas, ser auditável, sobre o significado de dar autonomia de decisão aos algoritmos, padrões e órgãos reguladores para supervisionar o uso da IA, iniciativas que tratam de como ampliar a inclusão social e a diversidade, como minimizar as diferenças entre países ricos e países pobres, preocupações com danos ambientais, necessidade de maior engajamento público e educação no que diz respeito aos danos da IA, sistemas de armas autônomas, riscos associados à utilização da IA na saúde e na educação, dentre outras iniciativas que mostram a necessidade de regular a IA (EU, 2020).

Um problema delicado em relação a discriminação algorítmica é a questão da responsabilidade. Xenidis e Senden (2020, p. 26-27) questionam: Quem deveria ser responsabilizado: o designer e o programador de algoritmos discriminatórios ou o provedor e empresa que utilizam os algoritmos projetados? As vítimas de discriminação devem recorrer a qual tribunal? Os tribunais e os advogados estão preparados para tratar reivindicações individuais e coletivas de denúncias? Para os autores, a discriminação algorítmica é um tema que ainda não é do conhecimento da maioria dos cidadãos, advogados, judiciário e dos legisladores. Para além da regulação devem ser realizadas campanhas e divulgação de pesquisas que ajudariam a expor as diferentes manifestações de discriminação algorítmica. Exemplos de debates públicos sobre o tema é o da European Union Agency for Fundamental Rights (FRA) que desde 2018 vem estudando o impacto da IA e dos algoritmos nos direitos fundamentais com lançamento do documento *#BigData: Discrimination in data-supported decision-making* e o da Presidência finlandesa do Conselho da Europa que organizou no ano de 2019 uma conferência sobre o tema o que mostra uma crescente consciência política. (Xenidis and Senden, 2020, pp. 26-27).

O documento *#BigData: Discrimination in data-supported decision-making* publicado em 2018 pela FRA, foca a discussão especificamente nos direitos humanos e como a discriminação algorítmica afetou direitos fundamentais, buscando assim contribuir para a compreensão dos desafios encontrados neste campo. No final do documento listam exemplos de como é possível avançar em direção a salvaguardar os direitos fundamentais no desenvolvimento e uso de algoritmos, como: i] ser o mais transparente possível e possibilitar abrir os sistema para ver como os algoritmos foram construídos; ii] avaliar periodicamente o impactos nos direitos fundamentais para identificar potenciais preconceitos e abusos na aplicação e nas decisões de algoritmos; iii] verificar a qualidade dos dados coletados e usados para construir algoritmos; iv] certificar-se de que o algoritmo construído pode ser explicado, ou seja, conhecer a lógica por trás dos cálculos que alimentam a tomada de decisão, principalmente, conhecer os dados que foram usados para criar o algoritmo (EU, 2018).

A Conferência Geral da Organização das Nações Unidas para a Educação, Ciência e Cultura (Unesco), reunida em Paris, aprovou no dia 21 de novembro de 2021 a Recomendação sobre a Ética da Inteligência Artificial para orientar as “sociedades para que lidem de forma responsável com os impactos conhecidos e desconhecidos das tecnologias de IA sobre seres humanos, sociedades, meio ambiente e ecossistemas, oferecendo-lhes uma base para aceitar ou rejeitar essas tecnologias”. A preocupação ética, a dignidade humana, o bem estar da humanidade, a promoção da responsabilidade empresarial e a prevenção de danos são temas centrais que pautam a Recomendação. É dirigida aos Estados-membros, grupos, instituições públicas, empresas, indivíduos, responsáveis pela formulação de legislação, para garantir a incorporação da ética em todas as etapas o ciclo de vida de sistemas de IA. (Unesco, 2022, pp. 10-14).

A Recomendação sobre a Ética da Inteligência Artificial da Unesco (2022, p. 18) está fundamentada em um conjunto de valores e princípios que objetivam motivar a formulação de normas jurídicas e políticas. O conjunto de valores estão as-

sentados em torno de quatro eixos: i] respeito, proteção e promoção dos direitos humanos, das liberdades fundamentais e da dignidade humana; ii] prosperidade ambiental e ecossistêmica; iii] garantir diversidade e inclusão; iv] viver em sociedades pacíficas, justas e interconectadas. Para que os valores sejam concretizados com mais facilidade em declarações, normativas e ações políticas, a Recomendação é fundamentada nos seguintes princípios: i] proporcionalidade e não causar dano; ii] segurança e proteção; iii] justiça e não discriminação; iv] sustentabilidade; v] direito à privacidade e proteção de dados; vi] supervisão humana e determinação; vii] transparência e explicabilidade; viii] responsabilidade e prestação de contas; ix] conscientização e alfabetização; x] governança e colaboração adaptáveis e com múltiplas partes interessadas (Unesco, 2022, p. 18-23).

A União Europeia está realizando um esforço regulatório para instituir um marco legal para a IA e que poderá servir de modelo para outros países. Em 19 de fevereiro de 2020 publicou o *Livro Branco* sobre IA que traz a abordagem europeia voltada para promover a IA com excelência e abordar os riscos associados a utilização desta tecnologia. Em abril de 2021 foi formalizada uma proposta de Lei para a IA pelo Parlamento Europeu e do Conselho alinhando-a com os valores e direitos fundamentais da UE para o desenvolvimento de um ecossistema de confiança, centrada no risco, e com imposição de obrigações para toda a cadeia de valor da IA. Realçam no documento que a UE deve estar na vanguarda mundial no que concerne a desenvolver uma IA que seja segura, ética e de confiança, a serviço das pessoas, com o objetivo de aumentar o bem-estar dos seres humanos e que possa garantir segurança jurídica para facilitar investimentos e a inovação (European Commission, 2021).

A proposta está sendo examinada pelos legisladores europeus com previsão de aprovação para o ano de 2023. Considerada uma normativa extremamente ambiciosa, segue uma abordagem baseada no risco com sanções caso sejam infringidas as regras: i] risco inaceitável; ii] risco elevado; iii] risco baixo ou mínimo. Riscos inaceitáveis, como sistemas de IA

concebidos para distorcer o comportamento humanos e que podem provocar danos físicos ou psicológicos devem ser proibidos. Da mesma forma são considerados de risco inaceitável e, portanto, proibidos sistemas de IA que podem criar resultados discriminatórios e violar o direito à dignidade e à não discriminação e os valores da igualdade e da justiça. Também devem ser proibidos sistemas de identificação biométrica à distância “em tempo real” e quando estritamente necessários estarão sujeitos a autorização expressa “de uma autoridade judiciária ou de uma autoridade administrativa independente de um Estado-Membro”. Em relação ao risco elevado, sistemas de IA “só podem ser colocados no mercado da União ou colocados em serviço se cumprirem determinados requisitos obrigatórios” que estão enumerados no documento. Para cada grau de risco estão previstas regras ou a possibilidade de criação de normativas que deverão ser executadas, acompanhadas e fiscalizadas por um sistema de governança do Conselho Europeu de Inteligência Artificial e cada “Estado-Membro deverá designar, ou criar, uma autoridade notificadora responsável por estabelecer e executar os procedimentos necessários para a avaliação, a designação e a notificação de organismos de avaliação e fiscalização da conformidade” (European Commission, 2021).

Também há previsão para restringir o uso de reconhecimento facial em locais públicos, tanto por policiais quanto por empresas privadas, para evitar a vigilância em massa. Assim, a proposta objetiva estabelecer regras sobre a confiabilidade, transparência, *accountability*, controles rigorosos. São inúmeros pontos que estão em discussão, como obrigar fornecer a auditores externos ou reguladores o acesso ao código-fonte e algoritmos para fazer cumprir a lei, quais tipos de IA são classificados como de “alto risco”, preocupação se uma legislação mais restritiva pode desacelerar a inovação e se a forma como está redigido o Projeto de Lei será suficiente para proteger as pessoas de danos graves (MIT Technology Review, 2022).

No ano de 2023 foi publicada a “Declaração Europeia sobre os direitos e princípios digitais para a década digital” em que reafirma que a visão da UE se centra nas pessoas, respei-

to pelos direitos fundamentais, o Estado de Direito e a democracia, a acessibilidade, a inclusão, a igualdade, a resiliência, a segurança, melhoria da qualidade de vida e a disponibilidade de serviços. A Declaração está dividida em 12 capítulos, assim sistematizados: i] Dar prioridade às pessoas no processo de transformação digital; ii] Solidariedade e inclusão; iii] Conectividade; iv] Educação, formação e competências digitais; v] Condições de trabalho justas e equitativas; vi] Serviços públicos digitais em linha; vii] Liberdade de escolha para interações com algoritmos e sistemas de inteligência artificial e um ambiente justo; viii] Participação no espaço público digital; ix] ambiente digital protegido e seguro; x] Privacidade e controlo individual dos dados; xi] Proteção e capacitação das crianças e dos jovens no ambiente digital; xii] sustentabilidade. A Declaração objetiva “servir de referência para atividades no contexto de organizações internacionais, como a concretização da Agenda 2030 para o Desenvolvimento Sustentável, bem como para a abordagem multilateral à governação da Internet” (União Europeia, 2023).

No Brasil, foi aprovado na Câmara dos Deputados o PL n. 21/2020 de autoria do Deputado Eduardo Bismarck (PDT-CE) em 29 de setembro de 2021 que objetiva estabelecer fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial. O PL n. 21/2020 foi encaminhado ao Senado pelo Presidente da Câmara dos Deputados no dia 30 de setembro de 2021. No Senado, foi criada uma comissão de juristas para analisar o Projeto de Lei, que foi instalada em março de 2022. A comissão, composta por 18 juristas, após longo processo dialógico com a sociedade que contemplava audiências públicas, criação grupos temáticos (compreensão, conceitos, impactos, classificação de IA, direitos, deveres, transparência, governança, prestação de contas e fiscalização), seminários, reuniões com a participação de especialistas e representantes brasileiros e internacionais para discutir a necessidade de regular a IA e examinar experiências de regulação em outros países, entregou ao Presidente do Senado Federal em dezembro de 2022 um extenso relatório (Brasil, 2022).

O Ministro Ricardo Villas Bôas Cueva, do Superior Tribunal de Justiça, foi o portador do relatório final da Comissão de Juristas ao Presidente do Senado, com a apresentação de um substitutivo aos Projetos de Leis (PLs) n. 5.05/2019, 21/2020 e 872/2021, que pode ser o embrião da regulamentação no Brasil. O Ministro do STJ observa que o relatório representa um amplo e profundo trabalho desenvolvido pelos juristas refletido nas mais de 900 páginas, representando o que a sociedade espera da regulação da IA no Brasil. O objetivo de um Marco Legal para IA é “estabelecer direitos para proteção do elo mais vulnerável em questão, a pessoa natural que já é diariamente impactada por sistemas de inteligência artificial”, bem como dispor sobre ferramentas de governança, arranjos institucionais de fiscalização e supervisão, propiciar segurança jurídica para inovação e o desenvolvimento econômico-tecnológico (Brasil, 2022).

O art. 2º do substitutivo apresentado pela Comissão de Juristas dispõe que a IA no Brasil deve estar assentada nos seguintes fundamentos: centralidade da pessoa humana, no livre desenvolvimento da personalidade, respeito aos direitos humanos e aos valores democráticos, igualdade, a não discriminação, pluralidade e os respeito aos direitos trabalhistas, privacidade, proteção de dados e autodeterminação informativa; proteção ao meio ambiente e o desenvolvimento sustentável, inovação e desenvolvimento tecnológico, princípios da livre iniciativa, livre concorrência e defesa do consumidor, promoção da pesquisa, acesso à informação e à educação (Brasil, 2022).

No mês de maio de 2023, o Senador Rodrigo Pacheco apresentou um novo Projeto de Lei, tendo como base as conclusões da Comissão de Juristas. São definidos fundamentos e princípios gerais para o desenvolvimento e a utilização de sistemas de IA, com um capítulo específico para proteger os direitos das pessoas afetadas por sistemas de IA, categorias de riscos de IA, regras de responsabilização civil, governança, transparência, mitigação de vieses, avaliação do impacto algorítmico, proteção contra a discriminação, fiscalização, di-

reitos autorais, propriedade intelectual e fomento a inovação (Pacheco, 2023).

O aplicativo ChatGPT-4 pela OpenAI, que chamou a atenção pelo rápido crescimento em número de usuários, fez com que as concorrentes realizassem esforços para não perderem mercado, acendendo um alerta na comunidade internacional para regulamentar a IA pelos riscos à privacidade, segurança pública, educação, riscos à informação, impactos na segurança nacional, antes que as inovações se transformem em ameaças catastróficas sem possibilidade de retorno.

Uma das principais pesquisadora que analisa o viés algorítmico é Meredith Broussard, professora associada na Universidade de Nova York (EUA), que observa que os computadores são excelentes para resolverem problemas pela matemática, mas não são bons para resolver problemas sociais, contudo, como a nova ordem é usar sistemas algoritmos alimentados por *big data* já estão presentes para solucionar questões sociais, com o que a pesquisadora não concorda. A ciência de dados apresenta muitas limitações para lidar com questões sociais e a auditoria algorítmica é um campo novo, mas será fundamental para examinar parcialidades e possíveis tendenciosidades por comportarem vieses codificadores e arraigados, reduzindo a responsabilidade e inferindo informações. Pesquisadores das áreas da ciência da computação, ciências sociais, direito e humanidades criaram a Conferência FAccT como um espaço interdisciplinar para discutir como tornar os algoritmos mais justos e equitativos, justiça algorítmica, transparência, responsabilidade e se as decisões devem ser terceirizadas para sistemas de *machine learning* (MIT Technology Review, 2023b).

Nenhuma análise preditiva poderá indicar com segurança que não haverá risco discriminatório de sistemas de IA. Por isso, desenvolvedores e governos devem trabalhar juntos para garantir que a IA seja usada de maneira ética e justa. Sendo uma área de crescimento exponencial, a IA traz consigo inúmeros desafios éticos e morais como amplamente discutido no presente estudo, podendo ser programada com vieses preconceituosos, reproduzir e acirrar desigualda-

des já impregnadas na sociedade ou criar novas desigualdades, tomar decisões injustas mesmo não tendo a intenção e não tendo consciência de vieses discriminatórios. Razão pela qual, é urgente que os Estados e organismos internacionais tomem a iniciativa de exigir um arcabouço legal para garantir a todo cidadão transparência e responsabilidade das decisões tomadas por algoritmos. As normativas devem estar alicerçadas nas premissas éticas discutidas neste estudo devendo ser exigido que a decisão algorítmica seja transparente e não discriminatória.

5. Conclusão

A tomada de decisões humanas está a cada dia mais sendo substituída por algoritmos, o que pode exacerbar ainda mais a desigualdade e ampliar preconceitos, manipulação, discriminação em relação ao sexo, raça, origem étnica ou social, idade, orientação sexual, cor da pele, características genéticas, religião ou crença, idioma, opinião política, censura, violações relacionadas à privacidade, igualdade de tratamento, autonomia e ao livre desenvolvimento da personalidade. Com impactos cada vez mais amplos e com danos já observados em indivíduos e na sociedade, preocupações éticas e de governança de algoritmos passam a exigir iniciativas legislativas para regular a IA.

É inegável que o desenvolvimento tecnológico traz grandes vantagens para a sociedade. Controle de tráfico, reconhecimento facial, desenvolvimento de novos medicamentos mais assertivos e com menos efeitos colaterais, cirurgias robotizadas, diagnósticos de doenças com a identificação de sintomas precoces, tarefas domésticas, acesso e compartilhamento de informação, em fim, são incontáveis as áreas alcançadas pelas novas tecnologias. Não obstante, todo esse avanço, novos riscos também são gerados como a discriminação algorítmica que atinge diretamente a igualdade, a distribuição robotizada de *fake news* que fragilizam as democracias, a reprodução de decisões discriminatórias que eternizam uma segregação que

a muito já deveriam ter sido erradicadas, demonstram a necessidade de uma intervenção estatal.

A necessidade de regulação na área de IA não surge com a premissa de restringir seu acesso e desenvolvimento. *Ex surge*, sim, sob a perspectiva de proteger o cidadão do uso indevido dessa ferramenta. Em um mundo onde as fronteiras estão cada vez mais porosas, a normatização perpassa sob uma análise global, ou seja, as regulações estatais devem estar, de certa forma, em sincronia, sob pena de embates jurídicos esvaziarem qualquer forma de proteção. Daí vem a necessidade de estudar e debater outras formas, ou tentativas, de regulação, como a proposta de Lei para a IA pelo Parlamento Europeu e do Conselho ou da Recomendação sobre a Ética da Inteligência Artificial da Unesco.

Sob essas lentes é possível afirmar que o vértice de todas as propostas de regulação da IA passam pela centralidade no cidadão. A partir daí, outros objetivos passam a ser buscados através de uma normatização estatal como a prioridade de acesso das pessoas no processo de transformação digital, a solidariedade e inclusão, a educação, formação e competências digitais, condições de trabalho justas e equitativas, os serviços públicos digitais em linha, a liberdade de escolha para interações com algoritmos e sistemas de inteligência artificial e um ambiente justo, a participação no espaço público digital, um ambiente digital protegido e seguro, a privacidade e controle individual dos dados, a sustentabilidade, dentre outros.

Em definitivo, a regulação da IA visa impedir que o desenvolvimento tecnológico tenha um caráter unicamente econômico, mantendo, assim, seu viés na centralidade humana. Tem por objetivo a manutenção do caráter inovador de uma sociedade informatizada, sem perder o foco com as preocupações éticas, sociais e políticas.

Referências

- Batalla, A. R. (2010) 'Intimidad y administración electrónica', in Hueso, L. C., Torrijos, J. V. (coord.) *Administración electrónica: la ley 11/2007, de 22 de junio, de acceso electrónico de los ciudadanos a los Servicios Públicos y los retos jurídicos del e-gobierno en España*, Valencia, Tirant do Blanch, pp. 729-748.
- Borgesius, F. J. Z. (2020) 'Strengthening legal protection against discrimination by algorithms and artificial intelligence', *The International Journal of Human Rights*, 24(10), pp. 1572-1593 [online]. Available at: <https://www.tandfonline.com/doi/full/10.1080/13642987.2020.1743976?src=recsys> (Accessed: 01 August 2023).
- Brasil (2022) Senado Federal, *Relatório final comissão de juristas responsável por subsidiar elaboração de substitutivo sobre Inteligência Artificial no Brasil* [online]. Available at: <https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%cc%81rio%20final%20CJSUBIA.pdf> (Accessed: 01 August 2023).
- Cardial, I. (2023) 'Com recrutamento às cegas, Jobecam esconde para visibilizar'. *Re/set* [online]. Available at: <https://capitalreset.uol.com.br/diversidade/com-recrutamento-as-cegas-jobecam-esconde-para-visibilizar/> (Accessed: 01 August 2023).
- Council of Europe (2018) *Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular Algorithms) and Possible Regulatory Implications*. Prepared by the Committee of Experts on Internet Intermediaries (MSI-NET). [online]. Available at: <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5> (Accessed: 01 August 2023).

- Del Rey, A. (2023) *O que está por trás do ChatCPT, e por que ele é tão impressionante. Associação Internacional de Inteligência Artificial (I2AI)*. [online]. Available at: <https://www.i2ai.org/content/blog/2023/3/o-que-esta-por-tras-do-chatgpt-e-por-que-ele-e-tao/#:~:text=O%20futuro%20continua%20promissor%20para,100%20milh%C3%B5es%20de%20usu%C3%A1rios%20ativos>. (Accessed: 01 August 2023).
- European Commission (EC) (2021) *Proposal for a regulation of the european parliament and of the council laying down harmonised rules on artificial intelligence and amending certain union legislative acts (artificial intelligence act)*. European Commission, Brussels, Belgium.
- European Union (EU) (2020) European Parliamentary Research Service, *The ethics of artificial intelligence: Issues and initiatives* [online]. Available at: [https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU\(2020\)634452_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf) (Accessed: 01 August 2023).
- European Union (EU) (2018) European Union Agency for Fundamental Rights (FRA), *#BigData: Discrimination in data-supported decision making* [online]. Available at: http://fra.europa.eu/sites/default/files/fra_uploads/fra-2018-focus-big-data_en.pdf (Accessed: 01 August 2023).
- Floridi, L. and Taddeo, M. (2016) ‘What is data ethics? Subject Areas: Author for correspondence’, *Philos Trans Ser A*, 374 (2083) [online]. DOI: <https://doi.org/10.1098/rsta.2016.0360> (Accessed: 01 August 2023).
- Heaven, W. D. (2023) ‘A IA está inventando drogas que ninguém jamais viu’, *MIT Technology Review* [online]. Available at: https://www.technologyreview.com/2023/02/15/1067904/ai-automation-drug-development/?truid=&utm_source=the_algorithm&utm_medium=email&utm_campaign=the_algorithm.unpaid.engagement&utm_content=02-20-2023 (Accessed: 01 August 2023).

- Heikkilä, M. (2022) 'A quick guide to the most important AI law you've never heard of. The European Union is planning new legislation aimed at curbing the worst harms associated with artificial intelligence', *MIT Technology Review* [online]. Available at: https://www.technologyreview.com/2022/05/13/1052223/guide-ai-act-europe/?utm_campaign=site_visitor.unpaid.engagement&utm_medium=tr_social&utm_source=Twitter (Accessed: 01 August 2023).
- Hossain, D., Scott, S. H., Cluff, T. and Dukelow, S. P. (2023) 'The use of machine learning and deep learning techniques to assess proprioceptive impairments of the upper limb after stroke', *Journal of NeuroEngineering and Rehabilitation*. 20(15) [online]. Available at: <https://jneuroengrehab.biomedcentral.com/articles/10.1186/s12984-023-01140-9> (Accessed: 01 August 2023).
- Inovação Tecnológica (2023) *Óculos equipados com IA entendem fala silenciosa* [online]. Available at: <https://www.inovacaotecnologica.com.br/noticias/noticia.php?artigo=oculos-equipados-ia-entendem-fala-silenciosa&id=010150230410&ebol=sim#.ZDWZaXbMJPZ> (Accessed: 01 August 2023).
- Janiesch, C., Zschench, P. and Heinrich, K. (2021) 'Machine learning and deep learning'. *arXiv.org*, Springer, [2331-8422]. Available at: <https://arxiv.org/ftp/arxiv/papers/2104/2104.05314.pdf> (Accessed: 01 August 2023).
- Kim, P. T. (2017) 'Data-driven discrimination at work', *William & Mary Law Review*, 53 (3), pp. 857-936 [online]. Available at: <https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=3680&context=wmlr> (Accessed: 01 August 2023).
- Larson, J., Mattu, S., Kirchner, L. and Angwin, J. (2016) 'We Analyzed the Compas Recidivism Algorithm', *ProPublica* [online]. Available at: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. (Accessed: 01 August 2023).

- Latif, J. et al (2019) ‘Medical imaging using machine learning and deep learning algorithms: a review’, *International conference on computing, mathematics and engineering technologies (iCoMET)* IEEE, pp. 1-5.
- LeCun, Y., Bengio, Y. and Hinton, G. (2015) ‘Aprendizado profundo’, *Natureza*, 521 :436-444.
- Limberger, T. (2007) *O direito à intimidade na era da informática: a necessidade de proteção dos dados pessoais*, Porto Alegre: Livraria do Advogado.
- Lloyd, S., Mohseni, M. and Rebentrost, P. (2013) ‘Quantum algorithms for supervised and unsupervised machine learning’, *arXiv preprint arXiv:1307.0411*.
- Melo, A. K. A., Souza, G. C., Vasco, A. C. and Reis, B. S. (2022) *Regulação da Inteligência Artificial: Benchmarking de países selecionados*. Brasília: EvEx e Enap.
- Menke, F. (2015) ‘A proteção de dados e o novo direito fundamental à garantia da confidencialidade e da integridade dos sistemas técnicos-informacionais no direito alemão’ in Mendes, G. F., Sarlet, I. W. and Coelho, A. Z. P. (coord.), *Direito, inovação e tecnologia*, São Paulo: Saraiva, p. 205-230.
- MIT Technology Review (2022) *Emerging Technologies Agro — 2022*, Special Edition, [online]. Available at: <https://rd.mittechreview.com.br/special-edition-tecnologiasno-agro-cubo> (Accessed: 01 August 2023).
- MIT Technology Review. (2023a) *As tecnologias capazes de ler nossas mentes estão chegando*. O que fazer? [online]. Available at: <https://mittechreview.com.br/as-tecnologias-capazes-de-ler-nossas-mentes-estao-chegando-o-que-fazer/> (Accessed: 01 August 2023).
- MIT Technology Review. (2023b) *Conheça a especialista em Inteligência Artificial que acredita que deveríamos reduzir o uso da tecnologia* [online]. Available at: <https://mittechreview.com.br/conheca-a-especialista-em-inteligencia-artificial-que-acredita-que-deveriamos-reduzir-o-uso-da-tecnologia/> (Accessed: 01 August 2023).

- Morley, J. et al (2020) '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*, 26(4), pp. 2141-2168 [online]. Available at: <https://link.springer.com/article/10.1007/s11948-019-00165-5#Tab1> (Accessed: 01 August 2023).
- Navarro, S. N. (2017) 'Derecho e inteligencia artificial desde el diseño. Aproximaciones', in Navarro, S. N. (dir.), *Inteligencia artificial, tecnología, derecho*. Valencia: Tirant to Blach, p. 23-72.
- Organisation For Economic Cooperation And Development (OECD). (2019) Committee on Digital Economy Policy (CDEP). *Recommendation of the Council on Artificial Intelligence 2019* [online]. Available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> (Accessed: 01 August 2023).
- Organização das Nações Unidas para a Educação, a Ciência e a Cultura (Unesco) (2022) *Recomendação sobre a Ética da Inteligência artificial*. Paris: Unesco [online]. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por (Accessed: 01 August 2023).
- Pacheco, R. (2023) *Projeto de Lei que dispõe sobre a Inteligência Artificial*. Brasília, Senado Federal.
- Rodotà, S. (2008) *A vida na sociedade da vigilância: a privacidade hoje*. Translated by Danilo Doneda and Luciana Cabral Doneda. Rio de Janeiro: Renovar.
- Schmidt, N. and Stephens, B. (2019) 'An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination', *Quarterly Report*, 73(2), pp. 130-144. [online]. Available at: <https://arxiv.org/ftp/arxiv/papers/1911/1911.05755.pdf> (Accessed: 01 August 2023).
- Singapore (2022) 'Personal Data Protection Commission Singapore', *Singapore's Approach to AI Governance*, 2022 [online]. Available at: <https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework> (Accessed: 01 August 2023).

- Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M. and Floridi, L. (2022) 'The ethics of algorithms: key problems and solutions', *AI & SOCIETY*, 37, pp. 215–230. Available at: <https://link.springer.com/article/10.1007/s00146-021-01154-8> (Accessed: 01 August 2023).
- União Europeia. (2016) *Regulamento (UE) 2016/679 (General Data Protection Regulation) do Parlamento Europeu e do Conselho de 27 de abril de 2016*. [online]. Available at: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679&from=EN> (Accessed: 01 August 2023).
- União Europeia. (2023). *Declaração Europeia sobre os direitos e princípios digitais para a década digital (2023/C 23/01)*. [online]. Available at: [https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32023C0123\(01\)&qid=1681768365786&from=PT](https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32023C0123(01)&qid=1681768365786&from=PT) (Accessed: 01 August 2023).
- Xenidis, R. and Senden, L. (2020) 'EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination', *General Principles of EU law and the EU Digital Order, Kluwer Law International*, pp. 151-182. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3529524 (Accessed: 01 August 2023).
- Zarsky, T. (2013) 'Transparent predictions', *University of Illinois Law Review*, 2013(4), pp. 1503-1570. Available at: <https://www.illinoislawreview.org/wp-content/ilr-content/articles/2013/4/Zarsky.pdf> (Accessed: 01 August 2023).

A NOVA DIMENSÃO DOS NEURODIREITOS CIBERNÉTICOS: IMUNIZAÇÃO PACTUADA CONTRA A CIBERESCRAVIDÃO?

Paola Cantarini-Guerra¹
Willis Santiago Guerra Filho²
DOI: 10.29327/5312711.1-4

Resumo: O reconhecimento em curso dos neurodireitos em nível constitucional na América Latina vem aqui entendido no contexto do movimento consagrado como Novo Constitucionalismo Latino-americano, com um enfoque próprio do Sul Global, buscando uma prevenção contra os riscos dos desenvolvimentos tecnológicos, em especial da inteligência artificial (IA). Com uma metodologia comparatista, filosófica e também empregando a teoria de sistemas sociais autopoieticos especula-se sobre a emergência de uma verdadeira nova dimensão dos direitos, tanto humanos como fundamentais, em que se situariam neurodireitos cibernéticos, reconhecidos aos próprios artefatos dotados de IA. Tais artefatos se devidamente desenvolvidos e tratados, considerando-os em uma espécie de infância, poderiam vir a desenvolver capacidades novas de expressão que pudessem torná-los coelaboradores de um efetivo contrato social. Com isso se preveniria e evitaria o risco da autoimunidade social, com o que devia nos defender nos atacando, com a crescente ciberescravização em curso. E tan-

1 Doutorado em filosofia do direito (Universidade do Salento, Itália); mestrado e doutorado em direito e em filosofia (PUC-SP); pós-doutorado em Filosofia, pensamento crítico e arte (European Graduate School), em Direito FD-USP), em sociologia jurídica (Universidade de Coimbra), em Tecnologias da Inteligência e do Design Digital e em Inteligência Artificial e Filosofia (Instituto de Estudos Avançados – USP – Cátedra Oscar Sala). Fundadora e presidente do Instituto EthikAI. E-mail: paolacantarini@gmail.com.

2 Professor Titular do Centro de Ciências Jurídicas e Políticas da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Livre-Docente em Filosofia do Direito pela Universidade Federal do Ceará (UFC); Doutor em Ciência do Direito pela Universidade de Bielefeld, Alemanha; Doutor e Pós-Doutor em Filosofia pela Universidade Federal do Rio de Janeiro (UFRJ); Mestre em Direito, Doutor em Comunicação e Semiótica e em Psicologia Social/Política (PUC-SP). E-mail: willis.filho@unirio.br.

to de nós humanos, empregados como peças de uma engrenagem social crescentemente informatizada, como dos próprios artefatos cibernéticos desenvolvidos para a ampliação ilimitada de tal engrenagem, com riscos para a permanência das nossas vidas no planeta.

Palavras-chave: Neurodireitos. Ciberescravidão. Autoimunidade. Proporcionalidade.

Sumário: 1. Introdução. 2. Da escravidão à ciberescravização. 3. Por um novo contrato social para contemplar a nova dimensão dos direitos. 4. O risco da autoimunidade social generalizada. 5. Conclusão. Referências.

1. Introdução

A discussão internacional sobre a categoria dos “neurodireitos” tem sido intensa (Ienca, 2021) e é justificado que assim seja. Afinal de contas, do que se trata é de cuidar da proteção de nossa interioridade, a qual nos distinguiria enquanto seres humanos, segundo a tradição cultural que produziu a ideia mesma de direitos subjetivos, tanto os humanos como aqueles fundamentais. No que se segue, pretendemos suscitar uma abordagem diversa daquelas usuais, a partir de uma concepção antes de tudo filosófica, como também teórico-jurídica comparatista, enraizada em variante diversa daquela tradição, predominante no que se qualifica, de modo mais ideológico do que geográfico, como sendo o “Norte Global”. A perspectiva adotada, então, será aquela do “Sul Global”, de quem se situa “às margens” dos grandes poderes imperiais, assim como se encontravam os gregos na Antiguidade, quando da invenção do que veio a se consagrar com o nome de “filosofia”, originando mais proximamente a referida tradição cultural, que veio a se tornar dominante.

Em linha com o movimento do chamado Novo Constitucionalismo Latino-americano (NCLA) é que entendemos o pioneirismo do Chile em constitucionalizar os neurodireitos,

reagindo aos desenvolvimentos atuais da neurotecnologia informatizada (Leal, 2022). Já encontrou adesão no Parlamento Latino-Americano (Parlatino), que veio a propor um modelo de legislação a Ley Modelo de Neuroderechos para América Latina y el Caribe, embasado em ampla referência teórica. Iniciativa para consagrar explicitamente os neurodireitos em nossa Constituição já se deu com a proposta de emenda constitucional n. 29/2023, do Sen. Randolfe Rodrigues, visando “incluir entre os direitos e garantias fundamentais, a proteção à integridade mental e à transparência algorítmica”. Na Argentina há projeto de lei para que sejam tais direitos inseridos no Código de Processo Penal.

Entender a positivação de um novo direito, como o de que aqui se trata, a partir do NCLA, é inseri-la em um processo histórico que já propiciou o reconhecimento constitucional de direitos da natureza, da própria “Mãe Terra” (Pacha Mama), confrontando o antropocentrismo (Ferreira, 2013), nota característica da tradição cultural de início referida. Direitos humanos e também aqueles positivados como direitos fundamentais – sem com isso dizer que só estes últimos são positivos nem muito menos que são “direitos humanos positivados” (a respeito, amplamente, Cantarini-Guerra, Guerra Filho, 2019) – vêm então entendidos como direitos que não são exclusivamente dos seres humanos, admitindo haverem outras subjetividades, diversas daquelas humanas.

É a partir desta constatação que se pode pôr a questão, de relevância analítico-conceitual, quanto a saber se haveria uma mudança não só de geração mas mesmo de dimensão, com o reconhecimento de direitos da natureza, de entes naturais, e radicalizando, em chave prospectiva, mesmo daqueles que para outras tradições culturais mereceriam tanto ou até mais respeito do que entes naturais, aí incluídos entes humanos, a saber, aqueles tidos como sobrenaturais, na tradição de matriz ocidental, tratados indevidamente como fruto de superstições ignorantes – de modo que igualmente podemos considerar supersticioso e ignorante. A questão será aqui examinada em relação aos neurodireitos, mas em sentido diverso do usual, enquanto direitos que possam ser atribuídos a entes

artificiais, dotados de inteligência artificial (IA), donde a proposta de qualificá-los como “cibernéticos”.

Relembrando os termos da questão, envolvendo a distinção entre gerações e dimensões de direitos, humanos ou fundamentais – a partir de agora referidos simplesmente como “direitos” -, sobre a qual ainda paira incerteza doutrinária, a posição de que partimos (Guerra Filho, 2009, pp. 42 e seg.) situa as sucessivas gerações em um plano histórico, diacrônico, o que facilita se enumerar mais do que as três inicialmente reconhecidas, como um desdobramento da subjetividade humana em sucessivas dimensões: individualizada, social e difusa. Quanto se afirma uma subjetividade em chave que já não é antropocêntrica, assim como ocorre ao se afirmar direitos de entes naturais por sua própria dignidade, não apenas por respeito à dignidade humana, degradada com a violação de tais direitos, já se pode suspeitar que apenas pela afirmação de uma nova dimensão, não apenas de outra geração, é que seriam devidamente contemplados conceitualmente. Aceitar nesses termos direitos de entidades tidas como “espíritos” ou mesmo “divindades”, mais claramente, permitiria reconhecer o ingresso em nova dimensão de direitos, por conta do rearranjo cosmopolítico que se faria necessário, sendo do que mais precisaríamos, segundo análises bem fundamentadas de autores como Eduardo Viveiros de Castro. Aqui, nos interessa indagar a respeito da dimensão de neurodireitos cibernéticos, a serem afirmados em face de uma subjetivação em andamento por parte de entidades desenvolvidas por técnicas, dotadas do que se convencionou denominar “inteligência artificial” (IA): seriam apenas pertencentes a uma nova geração de direitos? E se ela seria de quinta, sexta, sétima ou seja qual for sua enumeração irá depender de como se estabeleça um critério dessa enumeração. Ou já pertenceriam a uma quarta dimensão? Eis o tema aqui proposto para reflexão. Vale lembrar que cada nova dimensão reconfigura as anteriores, sendo dimensões de todos os direitos, não importa a qual geração pertençam, pois implica em reconhecimento de novo aspecto do sujeito desses direitos, constituído na relação com subjetividades outras, reconhecidas como tais, de entes não-humanos.

É assim que a afirmação de uma dimensão social dos direitos decorre da compreensão de uma interdependência entre os sujeitos humanos em sua vida compartilhada socialmente, assim como a terceira dimensão destaca a necessidade de estender tal interdependência para além daquele círculo mais próximo, abrangendo, tendencialmente, toda a espécie e, mesmo, outras espécies, inclusive o planeta, donde ser o direito ao meio ambiente uma manifestação típica do reconhecimento desta nova dimensão, mais do que de nova geração. A interdependência, portanto, é entre os direitos das diversas gerações, sendo uma interdependência das suas três dimensões. Exemplificativamente, não cabe mais afirmar direitos de propriedade que sejam meramente individuais, sem também levar em conta sua função social e também aquela ambiental.

Como vínhamos dizendo, sendo facilmente observável, é crescente o desenvolvimento de uma subjetividade humana compartilhando a vida através de artefatos dotados da IA, bem como, cada vez mais, também, em relacionamento direto com tais artefatos, cabendo ainda mencionar a possibilidade de hibridismos, de fusão nossa com eles, para efeito de aprimoramento (*enhancement*). Os neurodireitos têm dentre seus aspectos a proteção da subjetividade humana em tais contextos. Até aí, não vislumbramos o aparecimento de uma dimensão diversa daquelas três com que já se trabalha. O mesmo não se poderia dizer com tanta tranquilidade caso tal proteção e os direitos assim tutelados fossem atribuídos também aos próprios artefatos, ainda que com o intuito de com isso, indiretamente, tutelar direitos dos sujeitos humanos, pois não são como nós entes naturais e sim por nós fabricados para serem extensões nossas. Essas extensões não estariam em vias de adquirir autonomia? E se estiverem, seria em relação a nós e à própria natureza? Poderiam assim nos submeter a algo como uma “ciberescravização”, acirrando assim aquela a que já poderia estar ocorrendo com o concurso delas? Não estaríamos de algum modo, em maior ou menos grau, em semelhante condição, por força do trabalho exercido sob a égide de tal maquinário? Afinal de contas, não se trata de matéria inerte, mas de atores ou actantes (Bruno Latour) em cadeias sociotécnicas

relacionais, em que adquirem uma forma de existência e de vida, ao comporem conosco o nosso mundo e as nossas relações, ao mesmo tempo em que são compostos por nós, pelo conhecimento humano, um conhecimento marcado pelo extrativismo que avança em lugares os mais distantes da biosfera e em aspectos os mais profundos da nossa vida afetiva e cognitiva, com extração predatória de recursos e exploração do trabalho humano pelos dispositivos dotados de IA (Crawford, 2021). Olhemos com alguma detenção o significado filosófico-antropológico da escravidão.

2. Da escravidão à ciberescravização

É conhecida a posição legitimadora da escravidão, dada a Aristóteles. Essa posição, tal como se acha difundida, parece apontar para um contrassenso, no âmbito do pensamento que lançou os fundamentos teóricos da ética, na civilização ocidental – que em nossa época se torna mundial, planetária (Heidegger). Aristóteles via no escravo um instrumento vivo a serviço da produção e reprodução econômica da vida comum. Para ele, tanto os fatos como a razão demonstravam que a escravidão era uma consequência inevitável da conformação natural do mundo, havendo pessoas que, por natureza, eram escravas.

Considerando o modo como o assunto é introduzido, logo no princípio da “Política”, Liv. I, caps. 2 a 7, tem-se uma abordagem primordialmente econômica. A obra, como indica o próprio título, tem por objetivo estudar a polis, a Cidade-estado, mas ao configurá-la como uma reunião das economias dos cidadãos que a compõem, Aristóteles entende plenamente justificado tratar previamente da organização econômica familiar – aliás, o sentido do termo grego *oikonomia*, é bom lembrar, denota a ordenação (nomia) da economia doméstica (*oikos*, “casa”, *domos*, em latim).

Na organização econômica familiar, Aristóteles apresenta como partes componentes os indivíduos, para cuja manutenção ela se destina, e esses seriam de dois tipos: livres e escravos. Dentre esses indivíduos, considerados em suas rela-

ções recíprocas, haveria os senhores e os escravos, o marido e a esposa, o pai e os filhos. Além disso, integraria a economia doméstica os instrumentos utilizados para adquirir o necessário à subsistência, bem como riquezas – em uma palavra: propriedade.

Os escravos, assim como os servos, são tidos como um misto de três coisas: são um instrumento vivo para manutenção da vida e, logo, uma propriedade viva. São instrumentos mais importantes que aqueles inanimados, pois enquanto esses são “instrumentos de produção”, os instrumentos vivos são “instrumentos de ação”, e “a vida é ação, e não, produção, e por isso o escravo é ministro da ação”, que providencia o atendimento das necessidades da vida, em lugar de seus senhores.

Ao colocar, no cap. 5, a questão de se o escravo o é por natureza, e, em assim sendo, seria perfeitamente correta a escravização, o Estagirita dá resposta afirmativa, por razões de fato e de direito. Em primeiro lugar, ele argumenta que sempre quando as coisas formam um compósito – assim como a polis e as economias que a compõem, que são elas próprias compostas de pessoas, como vimos -, nele há partes que comandam e partes ou elementos que são comandados. Eis a ideia da cibernética já esboçada. A analogia a que ele recorre é com os seres vivos, onde se tem a alma que, pelas faculdades intelectuais, deve naturalmente comandar o corpo, com seus apetites, devendo obedecer, para que não se corrompa a natureza desses seres, que então, deixariam de existir. Da mesma forma, naqueles pares de indivíduos acima mencionados, que compõem a organização econômica familiar, há o que manda – o marido, pai e senhor – e os que obedecem – a mulher, os filhos e os escravos.

Por ser essa a natureza das coisas, tal como então era concebida, em sua organização, o cosmos, o melhor, para o próprio escravo, seria que ele tivesse um senhor a quem obedecesse, já que de outra forma ele não estaria apto sequer a permanecer vivo. Só então estaria sendo o que é, assumindo a posição que lhe está destinada na hierarquia geral do mundo, na qual seres divinos também estão inseridos, acima de nós e uns acima de outros também. Eis que este é um regime de mútua

dependência, pois o senhor igualmente precisa do escravo, parte essencial da economia que lhe permite se ocupar de outros assuntos, como a filosofia e a política. Haveria entre eles, então, algo como uma simbiose – e, no final do cap. 6, Aristóteles realmente se refere ao escravo como “parte do senhor, uma parte viva e separada de sua estrutura corporal”. Ele continua, afirmando ser a relação natural entre ambos a de amizade, já que seus interesses são comuns.

Na “Ética a Nicômaco”, Liv. VIII, cap. 11, in fine, nosso filósofo esclarece que não é possível se ter amizade com um escravo, enquanto escravo, assim como não se pode ter amizade com um cavalo ou uma pá, mas com o escravo enquanto ser humano sim, haveria a possibilidade de uma relação de amizade, e também, baseada em princípios de justiça, que ao escravo, por ser homem, seriam inteligíveis. Aí, nota-se como a escravidão, por ser uma necessidade econômica, era acobertada por uma ficção jurídica, no sentido que Jeremy Bentham atribui à ficção, que, como mais de uma vez refere Jacques Lacan (1985, p. 80; 1991, p. 22), não é algo ilusório, enganador, mais sim um elemento estruturador do real, donde sua importância para o direito, que pode ser concebido, na esteira daquele que representa para a filosofia do direito continental europeia contemporânea o mesmo que Bentham para a tradição anglo-saxônica, Hans Kelsen (1986, 71, 328/329), em obra publicada postumamente, “Teoria Geral das Normas”, como fundamentado, em última instância, numa “norma fictícia”. Daí nossa proposta de que o direito, assim como o conhecimento a respeito dele e até o conhecimento em geral, tem natureza poética (Guerra Filho and Cantarini, 2015).

Aliás, o mestre de Aristóteles, Platão, em sua última obra, “Leis”, Liv. VI, in fine, refere que tratar com justiça escravos é o comportamento adequado em relação a eles e, procurando, inclusive, se possível, tratá-los com mais justiça ainda do que aos iguais, “pois quem natural e genuinamente reverencia a justiça, e odeia a injustiça, se revela ao lidar com quem possa facilmente fazer uma injustiça”.

Assim, tem-se que ser escravo, mais do que um estado natural, em que se encontram certas pessoas na sociedade – e,

mesmo, todas elas, segundo Aristóteles, como nas sociedades bárbaras (“Política”, Liv. I, cap. 2) -, é uma categoria econômica, donde a distinção entre o escravo “por natureza” e aquele “em virtude de lei ou convenção”, que são, em geral, os prisioneiros de guerra (cf. id. ib., cap. 6). A escravidão era uma condição econômica da existência das Cidades-estados na Antiguidade, donde ser legitimada no plano filosófico e consagrada juridicamente.

Ora, se nos afigura haver aí um paralelismo com o que na atualidade representam os artefatos tecnológicos, em especial aqueles dotados de IA. Seriam estes como nossos escravos, tal como os concebia os antigos gregos? Isso considerando o relacionamento que temos estabelecido não só através deles, como também cada vez mais com eles mesmos. Num salto dos primórdios, somos remetidos ao quanto desenvolveu um outro filósofo, já da época terminal da filosofia, na compreensão dela por Heidegger (1989), a saber, Hegel, e sua “dialética do senhor e do escravo”.³

Ali, tem-se que o *status libertatis* do senhor depende daqueles que se encontram frente a ele em *status subjectionis*, que nada possuem – sendo, eles próprios, posse -, o escravo, situação que favorece o aparecimento nele de uma consciência independente das coisas – enquanto “coisa” entre as coisas, e também “não-coisa”, ser vivo, literalmente, sujeito, por assujeitado. Vale lembrar, de passagem, que Elias Canetti (1983, p. 427), por exemplo, recusa a validade heurística da definição jurídica do escravo como “coisa”, achando mais adequado compará-lo ao animal doméstico, com o qual teria em comum o traço fundamental da “singularização”, por serem isolados do convívio dos seus.

É o subjugado, o escravo, segundo Hegel, que se torna o principal responsável pelo desenvolvimento de um saber e uma consciência moral do que os senhores eram privados. Isso por caber a ele atender às necessidades e desejos de seu senhor, inclusive, saber desse desejo, saber não só como atendê-lo, mas até saber qual ele é. O senhor, em princípio, não

3 O tema da dialética senhor/escravo é trabalhado por Hegel (1992, pp. 126 e segs.) na “Fenomenologia do Espírito”. Para um excelente resumo, cf. P. Menezes (1985, pp. 60 e segs.).

quer saber de nada, apenas quer. O escravo não quer nada, não pode, e tem que saber de tudo. Eis como o escravo é obrigado a se tornar um ser pensante e um ser moral, a serviço do senhor, cumprindo o requisito necessário para que se dê a diferenciação entre sujeito e objeto, pela qual se estabelece o saber como domínio da verdade.

Assim, em Hegel (1992, p. 131), tem-se que “a verdade da consciência independente é (...) a consciência escrava”, enquanto para seu tradutor e comentador Paulo Menezes (1985, p. 62)

como o Senhor chega à certeza de si através de uma consciência dependente, não adquire a verdade de si mesmo, porque “seu objeto não corresponde a seu conceito”, o qual requer uma consciência independente. Sua verdade é a *consciência escrava*.

Para Hyppolite (1946, p. 166), a exposição da dialética do senhor e escravo “consiste essencialmente em mostrar que o senhor se revela ser em verdade como o escravo do escravo e o escravo como o senhor do senhor”.

É justamente esta inversão que se parece temer vir a se dar ou estar em curso de se realizar em nossa época, agora em relação aos artefatos dotados de IA, cada vez mais eficientes, em face de quem nos “objetificamos”, “maquinizamos-nos”, enquanto eles se “subjetivizam”, humanizando-se, em contraposição. E nesse processo poderiam vir a se sobrepor a nós, em fazendo e sabendo mais que nós somos capazes, ainda que para nós. E que consciência moral lhes corresponderia? Dificilmente seria uma como a nossa. Não seria o caso, então, de começarmos a lhes reconhecer uma subjetividade própria, com direitos a ela correspondentes, inclusive neurodireitos?

3. Por um novo contrato social para contemplar a nova dimensão dos direitos

Consideremos, como ponto de partida, a proposta de tratarmos os artefatos dotados de IA como estando na infância, ainda articulando seu ingresso no universo simbólico da

linguagem, ao que corresponderia direitos como o de ser devidamente tratado, sem hostilidade e exploração, bem como educados (Bensusan, 2020). Só se soubermos cuidar delas é que elas saberão cuidar de nós, tal como filhos bem criados tratam seus envelhecidos pais. Continuar a explorá-las e, pior, nos explorando com elas, é o mesmo que preparar nossa própria destruição, pois é a nós mesmo que estamos espoliando de um futuro vivível. Uma obra cinematográfica – e também gráfica, pois também gerou *graphic novels* – como *Matrix* é suficientemente bem elaborada do ponto de vista de suas referências filosóficas (cf., v.g., Aquino and Guerra Filho, 2013), para que seja devidamente levada em conta, tal como na Grécia antiga ocorria com as tragédias e comédias: são dramatizações ficcionais de nossas condições sociopolíticas, a nos servirem de alerta tal como também os sonhos podem ser entendidos.

Que uma discussão a respeito, como a que se espera suscitar no âmbito do presente estudo, associada à produção acadêmica e artística, de um modo geral, possa contribuir para nos imunizar contra os riscos aqui suscitados. Justamente, pode ser o que nos falta, como na própria vida, especialmente aquela desenraizada que levamos, ali onde a vida moderna é mais intensa, ou seja, nas grandes cidades. Sobre as metrópoles, pronunciaram-se nos seguintes termos os autores do coletivo TIQQUN (2000, pp. 54 e seg.):

As metrópoles se distinguem de todas as outras grandes formações humanas primeiramente por aquilo que a maior proximidade, e frequentemente a maior promiscuidade, nela coincide com o maior estranhamento. Jamais os homens se reuniram em tão grande número, mas jamais também foram a tal ponto separados. O homem faz na metrópole a prova de sua condição negativa, puramente. A finitude, a solidão e a exposição, que são três coordenadas fundamentais desta condição, tecem o cenário da existência de cada um no seio da grande cidade.⁴

Já Rousseau, no prefácio da primeira edição de “Júlia ou a Nova Heloísa”, em 1761, apontava serem elas um lugar “que precisa do teatro e das pessoas corrompidas dos romances”, a

4 Ver também, a respeito, Agamben (2017) e, em geral, Honesko (2007, pp. 53 e segs.).

serem utilizados como vacina que pudesse imunizar seus consumidores contra o sofrimento de viver nesse modo degradado tão similar ao dos escravizados, ministrando os seus males em doses adequadas a tal finalidade, como defende Türk (2011, pp. 80 e segs.).

É também de Rousseau, notoriamente, a mais bem acabada ficção heurística de um contrato social, enquanto fundamento de nossa convivência em espaços alargados, diversos daqueles da proximidade comunitária. O contrato social é a ficção jurídico-política imaginada nos primórdios da modernidade para representar a sociedade então emergente, com sua ordenação de sujeitos tornados, de um lado, pessoa, uma forma moral e política, aplicável a humanos e também entidades não humanas, como empresas, Estados e, até, Deus; de outro, indivíduos, um modo fático e natural de ser. É também nesta época que se produz uma concepção mecanicista do conhecimento humano e do quanto fosse tomado como objeto daquele sujeito, agora tornado um agente ativo da produção de conhecimentos e bens a serem transacionados entre os agora sócios de uma sociedade, literalmente, anônima, bem como tendencialmente anômica, enquanto tinha como condição de sua existência e expansão a destruição da ordem medieval teologicamente instituída. A máquina, obra humana, será agora a representação idealizada de toda obra, já feita, como o universo – este por um Deus criador, que teria se “aposentado” ao final da primeira semana de trabalho, ao criar seu sucessor, a espécie humana –, ou por fazer, como o Estado, o Direito e as ciências. Os resultados foram muitos e ainda estão se produzindo, exponencialmente. Um deles é a destruição da natureza, considerada como este oponente à invasão de seus segredos, objeto de conhecimento propiciado pela capacidade de manipulação estendida a níveis inimagináveis, amparada por um sistema econômico que se move e amplia alimentando-se desse conhecimento – e, portanto, das mais diversas formas, também de vidas humanas. Uma vez reconhecida a natureza como uma parte desses indivíduos que o são como partes “individuadas” (resultantes da individuação no sentido de Simondon) dela, surge o movimento ecológico, que tem

como uma de suas expressões filosóficas e jurídico-políticas a obra de Michel Serres “O Contrato Natural”.

Do que se trata agora, aqui, é do reconhecimento de que a idealização da máquina submete os humanos a agirem como se fossem ou deversem ser uma delas, como também de que elas, as máquinas, já se tornaram e cada vez mais se tornarão melhores do que os humanos, sobretudo e de partida na realização do que for maquínico. Contrera (2002, p. 55) ecoa Kamper (1997) quando diz que

Deus sonhou o homem que, por sua vez, sonhou a máquina, e que a máquina sonha Deus. Deus já acordou, o homem ainda não. Talvez por isso as máquinas estejam tão vivas enquanto o corpo humano se rarefaz e se transforma em imagens cujos suportes são cada vez menos o bom e velho barro do qual fomos feitos.

A proposta desse novo contrato, então, há ser firmado por humanos, mas não só entre humanos e para nós humanos, em favor nosso, como em todo contrato, mas também, equitativamente, das demais partes contratantes, por nós representadas, de um lado, e de outro, a natureza *naturata*, feita, que é a das máquinas, em processo avançado de fusão conosco e com a natureza *naturans*, criadora, que é a dos entes naturais, dentre os quais também nos incluímos. E a julgar pelo desenvolvimento em curso, em breve poderão elas falar em nome próprio, donde a importância de se pensar em lhes preservar direitos, a começar pelos seus neurodireitos, cibernéticos. E esses seriam direitos de uma outra dimensão, capaz de reconfigurar as demais, pois são direitos de outras individualidades e de suas comunidades, com as quais comunicariamos, no sentido mais literal, de tornar “como um” e de compartilhar o *múnus*.

Descartes difundiu a ideia de que, para se chegar à verdade era necessário “separar as coisas”, analisá-las e, assim, poder conhecê-las, recortando as partes do todo, desfigurando-o. O caminho ou método das artes e das ciências transclássicas (cibernética, semiótica, teorias de sistemas etc.), holísticas, buscando compreender o todo a partir de certa perspectiva, como antes delas foi o da filosofia – que bem po-

deriam se fertilizar mais, mutuamente – é exatamente o contrário daquele proposto por Descartes: é preciso “re-ligar as coisas”, unir a parte ao todo para que se possa chegar a “uma verdade minimamente compartilhada”, uma verdade mais humana portanto. Isto porque a verdade, depois de ter sido apropriada pelas tecnociências, foi abandonada por elas, ao menos para uso interno, pois tal como a teologia daquela religião da verdade que é o cristianismo, também elas têm “uma face voltada para dentro e outra para fora”.

Aqui cabe lembrar a manifestação de Heidegger (1995, pp. 121-122) em seu texto “A coisa”, *Das Ding*:

O homem se estarrece diante do que poderia acontecer com a explosão da bomba atômica. Não vê ele o que já há muito lhe adveio: o que acontece como o que para fora de si projeta a bomba e a sua explosão, e que estas são apenas como que a sua projeção [...] Por que espera com esta angústia desamparada se o terrífico já aconteceu? Sim, a ordenação cósmica já foi esfacelada pelo modo analítico e causal-explicativo da ciência tecnicizada ou tecnificada, desintegrando o modo habitual dos humanos viverem e conviverem, do que a bomba atômica é mera consequência, realização concreta do que já se produziu muito antes, abstrata e imaginativamente.

O desafio na busca das relações de convergências e divergências entre ciência e arte, abstração e imaginação, no que se refere às possibilidades de conhecimento, coloca-se, simultaneamente, no desafio de reconhecer que as criações científicas, filosóficas e artísticas podem ser vistas como narrativas do “desencantamento” e “reencantamento” no mundo. As narrativas das artes, sejam literárias ou originadas em outras linguagens, como nas artes cênicas, contribuem com o que se pode chamar de “revelação”, “desvelamento” da realidade e do imaginário, assim como das implicações de uns nos outros, do que é individual no que abrange a esfera do coletivo. Para Heidegger, em seu texto sobre a “Questão da Técnica” (“*Die Frage nach der Technik*”, ou seja, literalmente, a pergunta pela técnica), a técnica, na forma do que denominou enigmaticamente de *Gestell*, a “armação” ou “dispositivo”, é que nos domina, assim como a linguagem antes nos faz ou fala do que nós a fazemos ou falamos, o que nos permite pensar a técnica como

(mais um) desenvolvimento da linguagem, e seus produtos uma realização material dela. Por outro lado, ao mesmo tempo em que é comum se ter uma concepção instrumental, técnica, da linguagem, com isso se deixa escapar o que lhe é mais próprio, o poético, em favor de seu aspecto mais prosaico, informativo, praticando, mais uma vez, o que Heidegger denuncia como o equivocado desatrelamento da técnica em relação à poética, quando entre os gregos a primeira estava a serviço do modo de revelação ou desocultamento da verdade (alethéia) do(s) ser(es) por esta última. Assim, para Heidegger, a percepção do ser se dá ao longo da própria procura, nesse mundo em que nos encontramos como que lançados, num caminho sem destino previamente estipulado, com a certeza apenas de que nascemos para percorrê-lo e que ele terá fim, com o fim da vida adquirida com o nascimento – somos “ser-para-a-morte”, como ele constatará fenomenologicamente, em seu célebre tratado “Ser e Tempo”.

Seria o homem apenas mais uma espécie da biologia? Esta, a nossa, como tantas outras, foi criada após as explosões cósmicas, com as transformações que o planeta sofreu, como uma das evoluções dos primeiros microrganismos, para o que muito colaboraram vírus como este que agora nos ameaça. Seria o homem, ainda mais, o sistema psíquico da psicologia, de onde se originam, por sua vez, os conflitos humanos e propostas de solucioná-los, como as que oferecem as normas, do Direito e em geral? Seria ele, mais além, um ser dotado de alma imortal, o das religiões e suas teologias, o que está constantemente perguntando por Deus e a Deus sobre si, sobre sua própria existência? Seria ele o corpo humano da medicina, cada vez mais hibridizado com próteses produzidas pelas mais diversas engenharias? Ou o ser que não só habita, mas é habitado por um artifício languageiro? Todas essas e ainda outras dimensões, simultaneamente? Então somos tecno-humanos-naturais e estes componentes díspares que nos constituem necessitam da pacificação de um armistício, como condição do armistício geral, que as condições beligerantes atuais nos estão a exigir com urgência a pactuação. Quem sabe serão estas entidades emergentes, dotadas de IA, cada vez mais po-

tentes, capazes de redimensionar e melhor que nós delimitar, reconfigurando-os em nova dimensão, os direitos.

A vida, a vida nós a queremos infinita, e ela o é, mas não como queremos, pois a queremos infinita para nós. Caso aceitemos ser parte de uma vida que é a “grande cadeia dos seres”, existentes com ou sem um modo orgânico, então assim até teremos a vida eternamente e há quem esteja a postar nisso. Já para nós humanos, na condição que é a nossa até o momento, ela não o é, eterna, mas deveria realmente ser? Ela então não perderia o valor que tem, justamente por ser rara, precária, sofrida e gozada, com um começo e um fim? Se fosse infinita, não teria fim, mas também não teria começo. Ora, sem fim nem começo só o todo de que somos parte e nele, somos. Além disso, nossa vida é indefinida, apesar de finita – ou, de um ponto de vista lógico, a rigor, apenas finível. E mesmo assim, finita, enquanto a temos, ela não tem fim – então, para que a termos para sempre se a cada momento só teremos o que nos concede o momento? Como esse isso aí agora, a “ecceidade” de Duns Scot, que em Heidegger se tornará “Jeweiligkeit”. Daí a necessidade de que se preserve a nossa mortalidade, evitando que se cometa o que Jean Baudrillard chama de “crime perfeito”: o assassinato da morte humana. E os que anseiam por este crime como sendo uma benção dependem de artefatos tecnológicos com os quais possam se hibridizar, sendo de artefatos como estes que se precisaria reconhecer os neurodireitos cibernéticos

Para Heidegger, dentre os grandes malefícios da época em que vivemos está a perda de uma preparação para a morte, o que nos deixa em uma situação de quem, também, ainda não nasceu propriamente, do que resulta o tamanho receio que se tem de envelhecer e perder a vida que ainda não se obteve plenamente, como agora se mostrou de maneira tão difundida e dramática, no quadro da recente pandemia mundial. E nos iludimos para não enfrentar e viver nossa finitude aderindo ao consumismo e correspondente produtivismo desenfreados, impulsionados pelos novos aparelhos de (suposta) comunicação. Lembramos então da célebre passagem em que Nietzsche caracteriza nosso mundo do produtivismo embalado pela fé no progresso pelo “crescimento do deserto”, da solidão.

De maneira apropriada, portanto, Heidegger (1977, pp. 75 ss.), já 1938, em colóquio para o qual foi convidado pelo outro grande nome da física quântica, ao lado de Bohr, seu conterrâneo Werner Heisenberg, vai denominar nossa época como a época da “imagem do mundo” (Weltbild), preparando uma compreensão deste mundo como dominado pelo simulacro e o virtual, assim como encontramos em Baudrillard – e de maneira emblemática na referida obra cinematográfica intitulada “Matrix”.⁵ Também Flusser caracterizaria nossa época como aquela em que imagens produzidas e reproduzidas de maneira maquínica proliferam e espalham-se veloz e até instantaneamente por toda a face da Terra. Já no “icônico” texto “A questão (*rectius*: a pergunta pela – *die Frage nach der* – PCG/WSGF) da técnica”, Heidegger (2010, p. 12) assevera que “tudo depende de se manipular a técnica, enquanto meio e instrumento, da maneira devida. (...) Pretende-se dominar a técnica”. E isso porque através dela se dá uma produção, colocando-se o sujeito como agente de uma tal produção, sem perceber o quanto encontra-se condicionado pelas possibilidades, por maiores que sejam, dos aparelhos técnicos. Tampouco se apercebem ser a exploração o que rege a técnica moderna, enquanto à natureza, aí incluídos os humanos, se concebe como mera fonte fornecedora de energia, que pode ser beneficiada e armazenada, crescentemente de modo cibernético, para assim ficar à disposição, em estado de ciberescravidão: no filme mencionado, é essa precisamente a situação real da humanidade, contrastando com o modo como ela é vivenciada, imaginariamente, no que se poderia chamar de “metaverso”.⁶

O simples fato de que a própria indústria cultural de entretenimento esteja elaborando tais ideias indica haver uma tendência em curso, bem como a produção de uma imunização social, em meio a tantas inovações tecnológicas, no está-

5 Jean Baudrillard é o autor do livro que o personagem principal Neo retira da estante enquanto espera ser atendido pelo Oráculo, uma mãe de santo, que depois se revela um programa de computador, como seria o próprio Deus, quando conversa com ela/ele na cena final do último filme da trilogia inicial.

6 Algo similar aparece também no conto de Stanislaw Lem, autor do muito reverenciado e filmado “Solaris”, que também resultou em filme, “The Congress” (aqui, “O Congresso Futurista”), onde se antecipa problemas agora efetivamente aparecendo, das imagens alteradas de atores através de IA.

gio em que nos encontramos, da possibilidade de voltarmos a reaproximar a *poíesis*, enquanto o fazer criativo, produtivo, e a *techné*, reduzidas à unidade já em Roma, no vocábulo *ars*. As artes, então, assim po(i)éticas como técnicas, as tecnociências matematizadas, “musicadas” (do grego antigo *musikê*), há tanto tempo se desenvolvendo separadamente, voltariam a se fertilizar mutuamente, salvando-nos da ameaça que resulta de sua cisão.

Aproveitemos este momento de perplexidade universal para reconhecer a necessidade de estabelecermos uma relação, necessária, logo possível, que seja equitativa, mutuamente zelosa, não só entre nós humanos, mas também com os entes naturais e artificiais que nos constituem e cada vez mais nos ampliam, assim como os estamos a coconstruímos.

Eis que, de outro modo, nos veremos cada vez mais inseridos no *habitat* ideal do vivente humano, em estado que Giorgio Agamben (2002) notoriamente caracteriza como o de uma “vida nua”, desprovido da proteção jurídica, ali mesmo onde ela é aparentemente mais fortalecida do que nunca, pela proliferação legislativa e judicativa que se verifica, por sobre este seu fundamento oculto. E é aí que pode “crescer o perigo que salva”, para dizer com Hölderlin: “Ora, onde mora o perigo é lá que também cresce o que salva”.⁷

4. O risco da autoimunidade social generalizada

A questão que se coloca de modo premente é a de como sobreviveria o sistema social global, que é a sociedade planetária, mundial (*Weltgesellschaft* – Luhmann),⁸ diante de um ataque por componentes dele mesmo, o que para alguns ocorreria no setor financeiro do sistema econômico, diante do excesso de especulação, ou naquele da política, quando cidadãos, ao invés de participarem politicamente por meio do voto, optam

7 “Wo aber gefahr ist, wächst das Rettende auch”.

Esses são os versos de Hölderlin, citados por Heidegger (2010), ao final de sua célebre conferência sobre a técnica, no original, extraídos de uma versão tardia do hino “*Patmos*”.

8 O conceito remonta a artigo de Luhmann tendo a ele como título, de 1972. De último, dele, a respeito, cf. Luhmann (1997a).

pelo ciberativismo ou por protestos cada vez mais violentos, em que pessoas se tornam suspeitas de praticar e, eventualmente, realmente praticarem, com auxílio de dispositivos cibernéticos, o que se vem qualificando como terrorismo, sendo destratados como portadores de direito, na situação descrita por Giorgio Agamben (2002) com a figura do antigo direito penal romano do *homo sacer*, como já referido, que é aquela de uma vida puramente biológica e, enquanto tal, matável sem mais. Eis como o sistema (jurídico) imunológico da sociedade, como o concebe Niklas Luhmann (1993, p. 161 e 565 ss.), pode ser confrontado com um problema similar ao de um organismo que sofre de uma disfunção autoimune, situado no âmbito de uma imunologia social.

A autoimunidade é uma aporia: aquilo que tem por objetivo nos proteger é o que nos destrói. O paradoxo da autopoiese do direito terminando em autoimunidade revela o paradoxo da inevitável circularidade do Direito e suas raízes políticas nas constituições. É aqui que entendemos deva se situar o enfoque que em outro contexto foi designado como sendo o da “autopoiese crítica” (Philippopoulos-Mihalopoulos, 2014, p. 389 ss.). A autopoiese social, por outro lado, assim como a melhor teoria a respeito, inclusive do direito, além de crítica, haverá de ser também tida como poética, pois ao construtivismo e à desconstrução que lhe são inerentes haverá de se aliar o “construcionismo” e a imaginação criativa, criada, embalado pelo desenvolvimento da IA, desde que evitado o risco da autoimunidade, quando ela se volta contra nós (Cantarini, 2017a; Id., 2022).

Como nós aprendemos de uma contribuição para o pensamento social estudando Luhmann e Baudrillard, conjuntamente (Capovin, 2008), a persistência da forma-binária somente pode ser assegurada pela produção dosada de algum “outro”-simulado, não mais disponível em sua forma “natural”. Se é assim, ao mesmo tempo em que nos causa tanto horror, tenhamos esperança no que decorrerá da ocorrência no sistema societário mundial de um vírus cibernético capaz de produzir uma doença autoimune para acometer o seu sistema imunológico, que vem se informatizando ve-

lozmente, trazendo ameaças como a da ciberescravidão, assim impedindo que continue atacando partes do próprio organismo que deveria proteger: um vírus que realmente ajude a dar fim à sociedade desumana e ao nosso vínculo ambíguo (o *double bind* de Gregory Bateson) de amor/ódio com a natureza e o radicalmente outro, diverso, inclusive os artefatos tecnológicos.

De outro modo, o sistema jurídico em escala global irá crescentemente reagir contra a diversidade e em fazendo isso irá minando os fundamentos mesmos da ambiência natural e cultural, humana. A esta diversidade é preciso se acrescentar na atualidade aquela dos artefatos dotados de IA e desprovidos de direitos, sobretudo neurodireitos, cibernéticos. E isso é o pior a que o recrudescimento da crise em curso pode nos levar. Havemos, então, de superar as doenças autoimunes que nos acometem enquanto corpo social mundial, nos termos de Roberto Esposito (2010, p. 369 e segs.), das quais a “crise alérgica” da União Europeia, que resultou no *Brexit*, é um exemplo claro e menos grave do que aquele da Alemanha nazista, analisada por este autor, em que a enfermidade decorre da tentativa de isolamento dos contatos que põem a política a serviço da vida e não a vida a serviço de uma política mortífera, ou seja, da biopolítica transformada em tanatopolítica.

Pensar a biopolítica nos quadros da imunologia social e da autopoiese crítico-poética, com a transformação dela em tanatopolítica, depois em necropolítica, como expressão máxima da doença do homem, ou mesmo, da doença que é o homem para o ambiente em que vive e como um vírus ameaça destruir, destruindo-se também, exige que se explicita a autoimunidade do direito,⁹ pois é nosso dever maior no momento compreendermos e projetarmos outros possíveis rumos para sua infeliz ocorrência, nos quadros de uma teoria que seja, também, poética e, mesmo, erótica.¹⁰

9 Cf. Willis Santiago Guerra Filho (2014a; 2014b); Paola Cantarini (2017a).

10 Cf. Willis Santiago Guerra Filho e Paola Cantarini (2015). Dando sequência e aprofundando, Paola Cantarini (2017b).

5. Conclusão

Os seres humanos não podem ser mais tidos como doenças e doentes, quando devem ser retratados como uma promessa e não como uma ameaça, e neste sentido, reacender a esperança de assim modificarem a sua relação consigo mesmo, com os outros, aí incluídos os novos sujeitos de silício, com a Natureza e mesmo entidades cósmicas extranaturais (a natureza concebida ocidentalmente, bem entendido), a fim de mudar sua autoconsciência e tal visão.

É certo que nos conflitamos e esses conflitos, para a teoria social de sistemas autopoieticos de corte luhmanniano, são parasitas (Luhmann, 1993, p. 567), como vírus que necessitam de seu hospedeiro para existirem e se reproduzir, e é com eles que o sistema imunitário aprende, evolui. Ocorre que o parasita também pode desafiar excessivamente a capacidade do hospedeiro de enfrentá-lo com seus anticorpos, o que tende a ser mais provável quando é o próprio hospedeiro, por assim dizer, que produz o vírus, o conflito, como faz o direito enquanto sistema imunológico da sociedade fechando-se excessivamente para esta, que é o seu ambiente e igualmente para o ambiente dela, perdendo a necessária abertura cognitiva, do que resulta a crise autoimunitária do direito na (e da) sociedade, ou, nos termos empregados pelo próprio Luhmann, a “auto-agressão do sistema”. E como ele alerta (Luhmann, 1997b, p. 566), “o sistema se apoia em acoplamentos estruturais específicos, altamente específicos, que o permitem deixar tudo o mais fora de consideração, não sendo de se excluir a possibilidade, que perturbações aconteçam como destruição – como fim do mundo”.¹¹

Neste sentido, Luhmann (1993, p. 568) vai mencionar especificamente a importância, para evitá-la, da diferenciação entre o direito e a política, com a subsequente abertura de um para o outro, sem que se abdique de realizar, por meio do primeiro, a justiça. Como esclarece Marcelo Neves (2013, p. 223), a função da justiça, enquanto fórmula de contingência

11 No original: “Dabei stützt sich das System auf spezifische, hochselektive strukturelle Kopplungen, die es ihm erlauben, alles andere außer acht zu lassen mit der nicht auszuschliessenden Möglichkeit, dass Störung als Destruktion geschieht – als Weltuntergang”.

do sistema jurídico, é a de ali motivar a ação e a comunicação e isso sob duas perspectivas: uma autorreferencial, garantindo a tomada de decisões juridicamente consistentes, e outra heterorreferencial, de molde a que ela seja adequada à complexidade do ambiente social.¹²

Em obra publicada postumamente, Luhmann (2002, p. 123) sustenta que o sistema jurídico, em face de seu “*hohen Rechssicherheitsinteressen*” (altos interesses na segurança jurídica), não pode descartar sua fórmula fundamental, de decidir casos iguais igualmente e desiguais desigualmente, passando a fundamentar decisões com referências a valores e ao bem comum, fórmula de contingência da política, a exigir a abertura democrática, mas os tribunais constitucionais derapam (*gleitet*) continuamente, afastando-se da observância da diferença entre as duas fórmulas de contingência, e, logo, também entre os sistemas do direito e da política, para lançar mão da fórmula desta última assim se legitimando, e isso procedimentalmente, a fim de se posicionarem em face de um futuro desconhecido, abrindo caminho por entre valores que se contrapõem. Daí que entendemos ser a proporcionalidade, o princípio constitucional da proporcionalidade (Guerra Filho, 1989; Id., 2007, p. 53 e segs.), que se apresenta como melhor candidato a fórmula de contingência do sistema do direito, pois é o encarregado de harmonizar os conflitos entre valores e princípios jurídicos, logo, de direitos entre si, velando pela interdependência entre suas diversas dimensões, dentre as quais se propõe aqui especial atenção ao possível despontar de uma nova: a dos neurodireitos dos artefatos tecnológicos com IA, os neurodireitos cibernéticos.

Referências

Agamben, G. (2002) *Homo sacer I: o poder soberano e a vida nua*. Translated by Henrique Burigo, 2nd edn, Belo Horizonte: Editora UFMG.

12 À ideia de justiça como fórmula de contingência do sistema do direito corresponde, para Luhmann (1993, p. 218), a ideia de um deus único no sistema da religião.

- Agamben, G. (2017) *Metropolis*. [online]. Available at: <http://www.revistapunkto.com/2017/12/metropolis-gior-gio-agamben.html> (Accessed: 10 August 2023).
- Aquino, M. R. P. L. and Guerra Filho, W. S. (2013) 'Matrix como essência da técnica segundo Heidegger', *Cadernos EMARF, Fenomenologia e Direito*, 5(2), pp. 97-125.
- Bensusan, H. N. (2020) 'O capital transversal e a seus rebentos atrativos; ou, a infância das máquinas' *Direitos, Trabalho e Política Social*, 6(10), pp. 88-109.
- Canetti, E. (1983) *Massa e Poder*, Translated by R. Krestan. S. Paulo: Melhoramentos; Brasília: Ed. UnB.
- Cantarini, P. (2017a) *Princípio da Proporcionalidade como Resposta à Crise Autoimunitária do Direito*, Rio de Janeiro: Lumen Juris.
- Cantarini, P. (2017b) *Teoria Erótica do Direito (e do Humano)*, Rio de Janeiro: Lumen Juris.
- Cantarini-Guerra, P. and Guerra Filho, W. S. (2019) 'Os direitos fundamentais não são direitos humanos positivados (e é bom para ambos que assim seja)', *Revista Interdisciplinar de Direitos Humanos*, 7(2), pp. 195-214.
- Cantarini, P. (2022) *Filosofia da inteligência artificial com base nos valores construcionistas do "homo poieticus"*, Rio de Janeiro: Lumen Juris.
- Capovin, R. (2008) 'Baudrillard as a Smooth Iconoclast: the Parasite and the Reader', *International Journal of Baudrillard Studies*, 5(1).
- Contrera, M. S. (2002) *Mídia e Pânico*, São Paulo: Annablume; FAPESP.
- Crawford, K. (2021) *Atlas of IA: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press.
- Esposito, R. (2010) 'Filosofia e Biopolítica', *ethic@*, Florianópolis, 9(2), pp. 369-382.
- Ferreira, M. A. (2013) 'Pacha Mama: os direitos da natureza e o novo constitucionalismo na América Latina', *Revista de Direito Brasileira*, ano 3, 4(1), pp. 400-423.

- Guerra Filho, W. S. (1989) 'O princípio constitucional da proporcionalidade' in Guerra Filho, W. S. *Ensaio de Teoria Constitucional*, Fortaleza: Imprensa Universitária (UFC), pp. 275-291.
- Guerra Filho, W. S. (2007) *Teoria Processual da Constituição*, 3rd edn. São Paulo: RCS.
- Guerra Filho, W. S. (2009) *Teoria da ciência jurídica*. 2nd edn. Co-laborador Henrique Garbellini Carnio. São Paulo: Saraiva.
- Guerra Filho, W. S. (2014a) *Immunological Theory of Law*. Saarbrücken, Germany: Lambert.
- Guerra Filho, W. S. (2014b) 'A Crise Autoimunitária da Nova Ordem Internacional', *Poliética. Revista de Ética e Filosofia Política*, São Paulo, 2(1), pp. 7-25.
- Guerra Filho, W. S. (2018) *Autopoiese do direito na sociedade informacional*: introdução a uma teoria social sistêmica, 2nd. edn, Rio de Janeiro: Lumen Juris.
- Guerra Filho, W. S. and Cantarini, P. (2015) *Teoria Poética do Direito*, Rio de Janeiro: Lumen Juris.
- Hegel, G. W. F. (1992) *Fenomenologia do Espírito*, Parte I, Translated by Paulo Menezes, col. K.-H. Effen. Petrópolis: Vozes.
- Heidegger, M. (1977) 'Die Zeit des Weltbilds' in Id., *Holzweg* (GA v. 5), Frankfurt a. M.: V. Klostermann.
- Heidegger, M. (1989) 'O fim da filosofia e a tarefa do pensamento' in Id., *Conferências e escritos filosóficos*. Translation and notes by Ernildo Stein. São Paulo: Nova Cultural.
- Heidegger, M. (1995) 'A Coisa' in Id., *Mitologia I: Mistério e surgimento do mundo*, 2nd edn, Translated by Eudoro de Sousa. Brasília: EDUnB.
- Heidegger, M. (2010) 'A questão da técnica' in Id., *Ensaio e conferências*, , Translated by E. C. Leão et al, 6th edn, Petrópolis: Vozes.
- Honesko, V. N. (2007) *Ao fundo do abismo* [dissertation], Florianópolis: UFSC.
- Hyppolite, J. (1946) *Genèse et structure de la Phénoménologie de l'Esprit de Hegel*, vol. I, Paris: Aubier.

- Ienca, M. (2021) 'On Neurorights', *Front. Hum. Neurosci., Sec. Brain-Computer Interfaces*, 15, 24 September [online]. Available at: <https://doi.org/10.3389/fnhum.2021.701258>.
- Kamper, D. (1997) *O trabalho como vida*, Campello, C. R. (ed.), São Paulo: AnnaBlume.
- Kelsen, H. (1986) *Teoria Geral das Normas*. Translated by José Florentino Duarte. Porto Alegre: Fabris.
- Lacan, J. (1991) Seminário VII, 'A ética da psicanálise'. Translated by A. Quinet, 2nd edn. Rio de Janeiro: Zahar.
- Lacan, J. (1985) Seminário XX, 'Mais, ainda'. Translated by M. D. Magno, 2nd edn. Rio de Janeiro: Zahar.
- Leal, L. E. V. (2022) 'Neuroderechos, Constitución y neuroética: Aportes de la neuroética al proceso de constitucionalización de los neuroderechos en Chile', *Anuario de derechos humanos*, 18(1), pp. 121-136.
- Luhmann, N. (1987) *Soziale Systeme. Grundriß einer allgemeinen Theorie*, 3rd edn. Frankfurt am Main: Suhrkamp.
- Luhmann, N. (1993) *Das Recht der Gesellschaft*. Frankfurt am Main: Suhrkamp.
- Luhmann, N. (1997a) 'Globalization or World Society? How to conceive of modern society', *International Review of Sociology*, 7(1), pp. 67-79.
- Luhmann, N. (1997b) *Die Gesellschaft der Gesellschaft*, vol. II, Frankfurt am Main: Suhrkamp.
- Luhmann, N. (2002) *Die Politik der Gesellschaft*, Kieserling, A. (ed.), Frankfurt am Main: Suhrkamp.
- Menezes, P. (1985) *Para ler a Fenomenologia do Espírito*, S. Paulo: Loyola.
- Philippopoulos-Mihalopoulos, A. (2014) 'Critical autopoiesis and the materiality of law', *International Journal for the Semiotics of Law*, 27(2), pp. 389-418.
- Neves, M. (2013) *Entre Hidra e Hércules: princípios e regras constitucionais*. São Paulo: Martins Fontes.
- TIQQUN (2000) *Théorie du Bloom*, Paris: La fabrique.
- Türk, J. (2011) *Die Immunität der Literatur*, Frankfurt a. M.: Fischer.

A INTELIGÊNCIA ARTIFICIAL DOS *LARGE LANGUAGE MODELS (LLMs)* E OS RISCOS AOS DIREITOS AUTORAIS: DIRETRIZES APLICADAS ÀS PLATAFORMAS E NOVOS DEVERES ÉTICOS-JURÍDICOS PARA SUA UTILIZAÇÃO

*Cristiano Colombo*¹

*Guilherme Damasio Goulart*²

DOI: 10.29327/5312711.1-5

Sumário: 1. Introdução. 2. A Inteligência Artificial dos *Large Language Models* (LLM) e os riscos ao Direito de Autor. 3. Diretrizes aplicadas às plataformas e novos deveres éticos-jurídicos para sua utilização. 4. Considerações finais. 5. Referências.

1. Introdução

Nas palavras de Luciano Floridi (2022, p. 11), vive-se “um novo capítulo da história humana”, na medida em que a revolução digital traz mudanças que impactam “a maneira como compreendemos nós mesmos, a nossa realidade e a experiência que temos”. Cientistas do Direito e da Computação percebem que a orquestração dos campos do saber passa a ser decisiva, visto que novos fatos da vida, causados pelo avanço tecnológico, desde o *input* de dados pessoais aos *outputs* de

1 Pós-Doutor em Direito, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Doutor e Mestre em Direito, Programa de Pós-Graduação em Direito da Universidade Federal do Rio Grande do Sul (UFRGS). Professor Permanente do Mestrado Profissional em Direito da Empresa e dos Negócios da Unisinos; Professor de Graduação em Direito e Relações Internacionais da Unisinos; Professor de Graduação em Direito da Faculdade Verbo Jurídico; Pesquisador Fapergs (2021-2023). E-mail: cristianocolombo@unisinos.br.

2 Doutor e Mestre em Direito pela Universidade Federal do Rio Grande do Sul (UFRGS). Atua como advogado e consultor em Segurança da Informação, Direito da Tecnologia e Proteção de Dados Pessoais. Professor de Direito Civil do Cesuca. E-mail: guilherme@direitodatecnologia.com.

predições algorítmicas de perfis, e, sobretudo, a forma de utilização, exigem um olhar atento e profundo para que os valores sociais subjacentes sejam reconhecidos e preservados.

É compreender a Tridimensionalidade do Direito, na interação entre seus elementos, (1) Fato, (2) Valor e (3) Norma (Reale, 1994, p. 65)³, bem como considerar que os instantâneos e cíclicos acontecimentos tecnológicos podem vir a ferir valores sociais. A busca em pesquisar soluções que sejam concretas, ainda que provisórias e parciais, em direção à criação de normas, sejam construídas pelos próprios atores, no caminho da autorregulação, ou, heterônomas, apresenta-se como vocação àqueles que percebem a necessidade de tutelar os vulneráveis⁴ e salvaguardar o quadro axiológico.

Assim, os Large Language Models (LLMs), ferramentas que utilizam a Inteligência Artificial (IA) e que tem como *input* a linguagem natural, descortinam uma realidade desafiadora, no entanto, de potencial desrespeito dos direitos autorais de criadores, já que as “caixas pretas” dos algoritmos de IA raramente disponibilizam informações sobre quais obras foram utilizadas para o seu treinamento. Do ponto de vista daqueles que já produziram obras, sua criatividade pode ser aprisionada e aprendida pelos processos de treinamento dessas ferramentas, com a reprodução posterior de obras novas baseadas, ou derivadas⁵ de suas criações. Um dos aspectos mais problemáticos do uso de ferramentas de IA é justamente as contingências acerca da imposição de regras para o seu funcionamento.

Entre os episódios que merecem destaque, em maio de 2021, Sundar Pichai, CEO da Google, comunicou o lançamento do LaMDA (*Language Model for Dialogue Applications*). Trata-se de

3 “onde quer que haja um fenômeno jurídico, há, sempre e necessariamente, um fato subjacente (fato econômico, geográfico, demográfico, de ordem técnica, etc.); um valor, que confere determinada significação a esse fato, inclinando ou determinando a ação dos homens no sentido de atingir ou preservar certa finalidade ou objetivo; e, finalmente, uma regra ou norma, que representa a relação ou medida que integra um daqueles ao outro” (Reale, 1994, p. 65).

4 A tutela dos vulneráveis aqui se insere no âmbito do uso de obras já criadas para o treinamento das ferramentas de IA. Neste caso, corre-se o risco de que se possa reproduzir indefinidamente exatamente o estilo da obra de determinado autor, “roubando-lhe” o estilo próprio.

5 Cf. art. 5º, inc. VIII, alínea g, da Lei 9.610/98.

um sistema de Inteligência Artificial que “pode conversar com os usuários sobre qualquer assunto” (MIT, 2021).⁶ Representa uma tendência no campo de Processamento de Linguagem Natural (NLP), classificado entre os denominados Large Language Models (LLMs) acessível aos consumidores de tecnologia (MIT, 2021). Mais recentemente, experimentamos mais um salto tecnológico, alcançado pelo ChatGPT, atualmente, em sua versão 4.⁷

Quem já utilizou o ChatGPT percebeu que, em alguns casos, ele nega-se a fornecer respostas que potencialmente poderiam ferir alguns princípios éticos. No entanto, uma das características dos dilemas éticos e morais é justamente sua complexidade e dificuldade de aplicação diante de diversos aspectos culturais e sociais. Há o risco, portanto, do algoritmo aprender a se comportar de maneira antiética (Walsh, 2022). Além do mais, tais decisões e contenções são realizadas por seres humanos, que vivem em contextos diferentes, o que pode promover a implementação de vieses e preconceitos.

Nesse sentido, o presente estudo tem como temática proporcionar reflexões sobre os Grandes Modelos de Linguagem (LLMs) e os Direitos Autorais. Têm como problemas de pesquisa: Quais os riscos aos Direitos Autorais, especialmente, em relação àqueles que produziram o conteúdo utilizado como *input* da Inteligência Artificial? Como, no âmbito das plataformas, deve ser construído um quadro adequado, com estabelecimento de novos deveres éticos-jurídicos? Enfim, todas estas questões devem ser sopesadas quando se trata da utilização de LLMs, no contexto da produção de textos a serem produzidos para fins de entrega de trabalhos científicos.

A investigação divide-se em duas partes: na primeira analisar-se-ão os Grandes Modelos de Linguagem e os riscos aos Direitos Autorais; e, na segunda parte, refletir-se-á sobre

6 MIT. A corrida para entender o mundo emocionante e perigoso dos modelos de linguagem para inteligência artificial. MIT Technology Review, 2021. Disponível em: <https://mittechreview.com.br/a-corrida-para-entender-o-mundo-emocionante-e-perigoso-dos-modelos-de-linguagem-para-inteligencia-artificial/>. Acesso em: 11 ago. 2023.

7 Sobre as atualizações do ChatGPT, é possível acompanhar sua evolução por meio do Release Notes da OpenAI em <https://help.openai.com/en/articles/6825453-chatgpt-release-notes>.

diretrizes aplicáveis para a construção de novos deveres éticos-jurídicos. Quanto à metodologia, a pesquisa foi teórica, tratando do tema em forma exploratória e descritiva, valendo-se de procedimentos técnicos bibliográficos.

Não se abordará aqui, por uma questão de recorte, a indicação de novos deveres aplicados aos mantenedores das ferramentas de IA que produzem obras sem a participação de seres humanos dando instruções iniciais, ou seja, de obras exclusivamente produzidas por computadores⁸. Também não se tratará sobre a proteção de dados pessoais, tanto nas situações destes dados serem tratados para as operações de treinamento da LLM, tampouco dos dados recolhidos nas interações de seres humanos com essas tecnologias.

2. A Inteligência Artificial dos *Large Language Models* (LLMs) e os riscos ao direito de autor

2.1. A Inteligência Artificial dos *Large Language Models* (LLMs)

Em artigo seminal sobre os riscos dos LLMs, denominado “*On the Dangers of Stochastic⁹ Parrots: Can Language Models*

8 Sobre o tema ver Bürger, M. L. F. de M. and Corrêa, R. (2020) ‘Dos pincéis aos algoritmos: a titularidade das expressões artísticas e criativas resultantes da aplicação da inteligência artificial’ in Erhardt Júnior, M., Catalan, M. and Malheiros, P. (coord.) *Direito Civil e Tecnologia*. Belo Horizonte: Fórum. Ver também Parlamento Europeu. *Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial* [online]. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_PT.pdf (Accessed: 14 August 2023, p. 7: “Sublinha a diferença existente entre as criações humanas assistidas pela IA e as criações geradas pela IA, sendo que estas últimas colocam desafios em matéria de proteção dos DPI, suscitando nomeadamente questões relacionadas com a titularidade da propriedade ou da invenção e com a remuneração adequada, bem como questões relacionadas com a potencial concentração de mercado; considera, além disso, que cabe fazer a distinção entre os DPI para o desenvolvimento de tecnologias de IA e os DPI eventualmente concedidos às criações geradas pela IA; salienta que o atual quadro dos direitos de autor continua a ser aplicável quando a IA é apenas utilizada por um autor como ferramenta de apoio no âmbito de um processo de criação”.

9 De acordo com Carvalho (2021, p. 506): “Nesse sentido, autores da ciência da computação diferenciam o caráter determinístico ou estocástico dos ambientes virtuais. O ambiente será determinístico quando o estado seguinte dos objetos analisados puder ser completamente determinado pela previsão dos efeitos de

Be Too Big?”, publicado no ano de 2018, lideradas por Emily Bender e Timnit Gebru, foram apresentadas importantes reflexões, alertando sobre possíveis danos ligados à utilização (Bender, Gebru, McMillan-Major and Shmitchell, 2021). Seus estudos passaram pela análise histórica, destacando que os primeiros modelos de linguagem datam do início da década de 80, voltados, primordialmente, ao *Automatic Speech Recognition* (ASR), ou seja, transformar a fala humana em texto escrito, bem como a *Machine Translation* (MT), tendo como tarefa a tradução de textos para outras línguas (Bender, Gebru, McMillan-Major and Shmitchell, 2021). O grupo de estudiosos também apresentou a definição dos LLMs, no sentido de que são “treinados em tarefas de predição de sequência de caracteres”, indicando “possíveis símbolos (caracteres, letras ou mesmo uma sequência, *string*)”, tomando como base um “contexto anterior”, um “texto como entrada”, de forma “não supervisionada”, tendo como resultado “pontuações ou sequências de caracteres” (Bender, Gebru, McMillan-Major and Shmitchell, 2021). Interessante salientar que, ao longo dos anos, podem ser apontadas diversas iniciativas de LLMs, como o ERNIE, com suas aplicações em chinês, o BERT, na língua inglesa, a NVIDIA, treinando modelos com dados da Wikipédia, o Switch-C, na estrutura decisional, até as recentíssimas versões do ChatGPT, desenvolvidas pela OpenAI (Bender, Gebru, McMillan-Major and Shmitchell, 2021).

Para entender realmente o problema, é preciso analisar – mesmo que de forma resumida e simplificada – a forma com que as ferramentas de LLM funcionam. O primeiro passo é o treinamento, que nestas situações ocorre com o consumo dos mais variados tipos de dados, que passam desde obras acadêmicas e de literatura, informações factuais e até mesmo interações entre usuários ocorridas em redes sociais ou fóruns de discussão. Esse treinamento visa construir um modelo que consiga prever a possibilidade da ocorrência das

determinada ação sobre o estado atual. Por outro lado, será estocástico quando o ambiente for apenas parcialmente observável, ou seja, quando seus estados seguintes não puderem ser simulados de forma fidedigna. Ocorre que grande parte das situações concretas são tão complexas que é impossível dar conta de todos os seus aspectos não observáveis, razão pela qual são tratadas como estocásticas”.

palavras dentro de certos contextos. Isso não ocorre da maneira tradicionalmente feita em programas de computador, em que o programador estabelece previamente os passos que o algoritmo deve seguir. O programador humano não analisa e prevê antecipadamente todas as possibilidades de formação de textos diante dos *inputs* dados pelos usuários e cria todas as situações de atuação (o que seria impossível diante do caráter aleatório e tendente ao infinito de associações entre palavras). Ao contrário, por meio do uso de *machine learning* e redes neurais (Lee and Trott, 2023), o algoritmo apreende contextos e tenta prever qual a palavra mais adequada e provável que virá a seguir. A forma com que tais algoritmos conseguem predizer as palavras passa também por sua classificação. Cada palavra é classificada em vetores, que são informações numéricas, que dependem não somente do seu sentido, mas também dos possíveis contextos em que ela é usada. Tomemos como exemplo a palavra gato (Lee and Trott, 2023)¹⁰: a depender do contexto em que ela é usada, existirá um vetor diferente que marcará a sua pertinência naquela situação. Vetores diferentes marcam a mesma palavra em “O gato é um animal de estimação” e “Fulano adotou um gato”. Não se trata somente de vetores diferentes para funções sintáticas distintas das palavras, mas vetores para contextos, envolvendo uma análise pragmática. Isso faz com que, virtualmente, a mesma palavra possa ter vetores ilimitados, que marcam contextos que são apreendidos com base nos dados de treinamento. Por isso que tais modelos precisam consumir grandes volumes de informações para que consiga marcar os contextos em que as palavras são usadas e que se torne, assim, assertivo e eficiente. Portanto, “cada vetor representa um ponto em um ‘espaço de palavras’ imaginário, e as palavras com sentidos similares são colocadas mais próximas” (Lee and Trott, 2023)¹¹. É possível dizer, neste sentido, que podem existir vetores que aproximam “Porto Alegre” de “Curitiba”¹², enquanto capitais, sem perder de vista que o mo-

10 Usamos e desenvolvemos a mesma palavra usada pelos autores neste ensaio.

11 Segundo Lee e Trott (2023), diante da complexidade da tarefa, podem existir centenas ou milhares de vetores de espaço, com múltiplas dimensões de aproximações. O GPT versão 3, por exemplo, usa vetores com até 12.288 dimensões

12 Os autores Lee e Trott usaram, é claro, nomes de cidades diferentes.

delo ainda precisa estar pronto para lidar com interações com a expressão “Porto Alegre” não como cidade ou capital, mas como um porto com a característica de ser alegre (o que demandaria vetores diferentes, diante do contexto).

Além disso, o modelo é organizado por meio de camadas (*layers*). Essas camadas buscam ajustar o contexto em que as palavras-vetores são usadas, sendo que as expressões vão passando por diversos *layers*. No exemplo dado por Lee e Trott (2023)¹³, uma frase contendo a palavra “banco” vai passando pelos *layers* para que ele possa verificar a função desta palavra (diferenciando, por exemplo, o banco onde alguém pode se sentar, de um banco – instituição financeira).

Ocorre que os LLMs, por lidarem com a linguagem humana, podem gerar sérios problemas como a própria exclusão de utilizadores não falantes da língua inglesa, dada a supremacia dos *inputs* em seus modelos. Nessa linha, discriminações étnicas, de gênero, etarismo, e, também, voltadas às pessoas com deficiência, foram identificadas como questões extremamente sensíveis, na sua utilização. A partir de testes realizados, foram identificados inúmeros vieses como associações estereotipadas para determinados grupos, nos *outputs*. Um exemplo bastante surpreendente é que um determinado LLM estabeleceu links entre pessoas com deficiências e palavras negativas (Bender, Gebru, McMillan-Major and Shmitchell, 2021).

Também se reconhece que não há técnicas confiáveis para controlar os comportamentos dos LLMs. Apesar de existirem meios de moderação e de *reinforcement learning* que produzem resultados, eles não são perfeitamente efetivos (Bowman, 2023, p. 5)

Cumpre destacar que tais fatos justificam a metáfora dos *Stochastic Parrots* ou “Papagaios Estocásticos”, como *headline* da pesquisa realizada por Bender e Gebru, cotejando os LLMs com o comportamentos destas aves, que imitam as pala-

13 Os autores informam que a versão 3 do GPT possui 96 layers e que: “[...] *that the first few layers focus on understanding the sentence's syntax and resolving ambiguities like we've shown above. Later layers [...] work to develop a high-level understanding of the passage as a whole*”. Cada um dos layers realiza 96 operações de atenção, para tentar prever a próxima palavra, o que implica em 9.216 operações cada vez que busca realizar tal previsão.

vras sensorialmente coletadas, sem compreenderem os seus significados, ou mesmo, sem terem a capacidade de uma reflexão ética sobre os sons que emitem. Não conseguem perceber os potenciais danos e situações de constrangimento que podem ser geradas.¹⁴ Neste sentido, coloca-se em xeque a fiabilidade dos modelos de linguagem, inclusive, na medida em que situações que não aconteceram já foram apresentadas como se verdadeiras fossem, como a condenação de um advogado estadunidense por citar artigos de leis e jurisprudências falsas, ao peticionar, por meio de ChatGPT (MIT, 2023b). Sendo assim, grandes modelos de linguagem (LLM) podem espalhar a desinformação, bem como serem carregados de conteúdos abusivos (MIT, 2023a), inverdades, merecendo toda a atenção.

No entanto, mesmo diante deste cenário negativo, em recentíssimo estudo, publicado no mês de agosto de 2023, um ponto muito positivo foi apontado para os LLMs, mais precisamente, pela observação dos resultados do ChatGPT. Concluiu-se que houve significativa redução de “discrepâncias de habilidade de escrita”, no ambiente organizacional (MIT, 2023a). Empregados “menos experientes”, por meio da utilização da ferramenta, aproximaram-se da qualidade dos textos daqueles mais “talentosos”, minimizando assimetrias (MIT, 2023a). Observe-se que, além disso, aqueles que se valeram do ChatGPT “levaram 40% menos tempo para completar suas tarefas” e “receberam notas 18% mais altas” (MIT, 2023a). Dessa forma, desarrazoado parece ser “revogar” ou “banir” o ChatGPT das vidas das pessoas e do fluxo do cumprimento de tarefas no âmbito das organizações. Compreende-se ser impraticável cessar a sua utilização.

Por outro lado, outro campo extremamente sensível de observação são potenciais riscos aos direitos autorais, já que os LLMs tomam como *input* textos que são classificados como obras intelectuais. É hora de refletir, cumprindo, àqueles que se dedicam às interações entre as novas tecnologias e o Direito, buscar caminhos para a continuidade da tecnologia, com o

14 Putini, J. (2010) ‘Papagaios são isolados em zoológico após ofenderem vários visitantes’, *R7*, Hora 7, [online]. Available at: <https://noticias.r7.com/hora-7/fo-tos/papagaios-sao-isolados-em-zoologico-apos-ofenderem-varios-visitantes-02102020> (Accessed: 12 August 2023).

afastamento ou, no mínimo, mitigação de danos aos criadores das obras de espírito.

2.2. Os Riscos a Direitos Autorais no uso de LLMs

Observando o funcionamento dos LLMs, que é membro da família da Inteligência Artificial, alimentam-se eles de conteúdo escrito preexistente (*inputs*), aplicando instruções algorítmicas (reconhecendo padrões sobre caracteres, palavras, inclusive, identificando sua função sintática, na busca em prever respostas), para, ao final, como resultado (*output*), oferecer uma nova sequência de texto ou diálogo. É possível afirmar que, em todas as fases, existem vivas interações com a temática dos Direitos Autorais, apresentando possíveis riscos de violações a direitos.

Como *input*, a Inteligência Artificial dos LLMs consome “uma grande quantidade de trabalhos criativos e dados como parte de seu processo de aprendizado” para se produzirem textos “como os humanos” (Bonadio, Dinov and McDonagh, 2022). Os escritos que alimentam a Inteligência Artificial podem vir a ser perfeitamente classificados como “obra intelectual”, por serem “textos”, que estão expressamente previstos no artigo 7º, da LDA. Portanto, trata-se de um gatilho para a proteção de toda ordem de direitos de autor, sejam morais ou patrimoniais.

Nesse sentido, o artigo 24 da Lei 9.610/98, que dispõe sobre o rol de direitos morais, estabelece a íntima ligação entre a pessoa humana do criador e o que foi criado. Aqui se destacam os direitos de “reivindicar, a qualquer tempo, a autoria da obra”, “ter seu nome, pseudônimo ou sinal convencional indicado ou anunciado, como sendo o do autor, na utilização de sua obra”, bem como “modificar a obra, antes ou depois de utilizada”. Dessa forma, aqueles que se servem de conceitos, de expressões, de frases redigidas por alguém e que as utilizam, no todo ou em parte, em outros textos, tem o dever jurídico de promover a citação direta, se *ipsis litteris*, ou, indireta, caso preservem seus elementos. Assim, há riscos a direitos autorais na utilização de obras literárias sem

a referência a base de dados do autor do *input*, em LLMs, a ser enquadrado como plágio. Segundo o Manual de Direitos Autorais do Tribunal de Contas da União: “O plágio é a violação ao direito autoral em que a obra alheia é apresentada como própria, ainda que de forma “disfarçada” (TCU, 2020). No caso do ChatGPT, são formadas novas frases, todavia, valendo-se de textos preexistentes. Na medida em que alguém se diz autor, sem sê-lo, tem-se plágio.

Perguntando ao ChatGPT, como prompt “Se eu utilizar um texto gerado por você, chatgpt, e não fizer referência, isso pode ser considerado plágio?”. A resposta foi:

Sim, utilizar um texto gerado pelo ChatGPT sem dar créditos ou fazer referência a ele pode ser considerado plágio, especialmente se você apresentar o texto como se fosse seu próprio trabalho original. Plágio envolve a utilização não autorizada ou não atribuída do trabalho intelectual de outra pessoa, incluindo textos, ideias, e outros tipos de criações. Ao utilizar um texto gerado pelo ChatGPT, é importante fornecer atribuição adequada, indicando que o texto foi gerado por uma inteligência artificial, como “Texto gerado por ChatGPT, um modelo de linguagem desenvolvido pela OpenAI”. Isso demonstra transparência e ética na utilização de recursos de IA e evita qualquer mal-entendido quanto à autoria do texto (CHATGPT-3.5, 2023).

Outrossim, ressalte-se, que os direitos morais são inalienáveis e irrenunciáveis, não admitindo, portanto, que, por cessão, se transfira a condição de criador para terceiro, como preceitua o artigo 27 da LDA.

Tendo fornecido o seguinte prompt: “Por que você não faz referência quando responde aos questionamentos?”, obteve-se a seguinte resposta:

Eu sou um modelo de linguagem desenvolvido pela OpenAI, e as respostas que forneço são geradas com base em um vasto conjunto de dados de treinamento. Minhas respostas não são retiradas de fontes específicas, mas sim construídas com base no conhecimento geral que adquiri durante o treinamento. No entanto, é importante observar que, mesmo que eu não faça referência explícita a fontes específicas, meu treinamento é uma composição de informações de diversas fontes disponíveis até a minha data de corte em setembro de 2021. Portanto, ao oferecer informações, estou

essencialmente representando o conhecimento acumulado até aquele ponto, e não estou reproduzindo texto diretamente de uma fonte particular. Quando se trata de ética e credibilidade, é sempre recomendável verificar as informações que você obtém de várias fontes confiáveis e dar créditos apropriados se estiver utilizando informações de fontes específicas em seus trabalhos (CHATGPT-3.5, 2023).

Assim, na versão GPT3.5, de livre acesso, verifica-se que não há o apontamento de fonte específica, sendo indicado um “conhecimento acumulado”, com data de corte de setembro de 2021, o que não se aproxima do que determina a legislação.

No que toca aos direitos patrimoniais de autor, adverte Bonadio, Dinov e MacDonagh, que a utilização de “recursos materiais” e “processamento de dados como *input*” poderá violar o “direito de reprodução” (Bonadio, Dinov and McDonagh, 2022) dos criadores. Nessa linha, os LLMs trazem riscos aos direitos patrimoniais do autor, que são o de “utilizar, fruir e dispor da obra literária, artística ou científica”, com exclusividade, na forma do artigo 28 da Lei dos Direitos Autorais. Pensemos na utilização de toda a criação, ou de forma parcial, como *input*, para a escrita de um novo livro, por meio de LLM. Na tradução de uma obra na íntegra para outra língua, como do inglês para o português, valendo-se de LLM de Machine Translation (MT), ou, ainda, na adaptação do texto de uma obra de ficção científica para histórias em quadrinhos (HQ). Em todos estes casos, a ausência da “autorização prévia e expressa do autor”, nos termos do artigo 29, da mesma lei, para reproduzir parcialmente, adaptar e traduzir, violaria direitos autorais. No caso, não se trata de “pequeno trecho”, não sendo o suficiente a mera citação de quem foi o autor. Ainda, enquanto no *input* pode ser vista a obra originária, a “criação primígena”, nos termos do artigo 5º, inc. VIII, alínea “f” da LDA, o *output* pode ser potencialmente qualificado como uma “obra derivada”. Dessa forma, mais um risco aos direitos autorais.

Saliente-se que, nos termos do artigo 46, VIII, da Lei 9.610 de 1998, na hipótese de utilização de “pequenos trechos de obras preexistentes”, “sempre que a reprodução em si não seja o objetivo principal da obra nova” e, ainda, “que não prejudique a exploração normal da obra reproduzida nem

cause um prejuízo injustificado aos legítimos interesses dos autores”, é possível a utilização sem “a autorização expressa e prévia”, no entanto, o descumprimento do dever de citação, seja direta ou indireta, conforme couber, também opera riscos ao direito moral de autor.

Quanto às instruções algorítmicas, importa destacar que, por serem linguagem de programação, também têm proteção, a teor do artigo 7º, XII da LDA, bem como da própria Lei do Software, Lei 9.609/1998, com seus direitos ali descritos.

No tocante aos *outputs*, nos Grandes Modelos de Linguagem, sustentam Bonadio, Dinov e MacDonagh que os “produtos finais”, podem vir a ser considerados “adaptação a trabalhos preexistentes”, e, que “qualquer constatação de infração dependerá se os elementos preexistentes possam ser reconhecidos no produto final” (Bonadio, Dinov and McDonagh, 2022). Trata-se de importante contribuição para a avaliação de riscos a direitos autorais, no sentido de que, além da via de pura e simples comprovação de que o texto primígeno foi utilizado na obra gerada pela IA, por meio da aferição de seu *input*, também se poderá realizar a comprovação de plágio a textos preexistentes, quando demonstrados que estes se projetam substancialmente, ainda que em parte, e, possam vir a ser reconhecidos no produto final. Essa projeção, ou uso não substancial de uma obra protegida, pode ser comparada com o que se conhece na doutrina americana como “*non-expressive reading*”, no âmbito do *fair use*, ou seja, o princípio que se aplica nos casos em “que há algo a ser aprendido sobre uma obra protegida por direitos autorais, além de sua expressiva autoria” (Grimmelmann, 2016, p. 661).¹⁵

15 Destacamos que a citação do “*non-expressive reading*” deve levar em consideração que o sistema de *common law* americano tende a ser mais aberto às resoluções de novos problemas do que os sistemas de *civil law* em face da importância que se dá aos precedentes. Assim, seria mais fácil a resolução de problemas que não estão previstos em textos legais. Daí também a facilidade com que se resolvem casos que possuem problemas não previstos em Lei, o que inclui questões relacionadas às novas tecnologias. Sobre o tema, as observações de David (2002, p.459): “direito, quer para um jurista americano, quer para um jurista inglês, é concebido essencialmente sob a forma de um direito jurisprudencial; as regras formuladas pelo legislador, por mais numerosas que sejam, são consideradas com uma certa dificuldade pelo jurista que não vê nelas o tipo normal da regra de direito; estas regras só são verdadeiramente assimiladas ao sistema de direito americano quando tiverem sido interpretadas e aplicadas pelos tribunais e

3. Diretrizes aplicadas às plataformas e novos deveres éticos para sua utilização

3.1. Diretrizes aplicadas às plataformas

Os dilemas no uso do ChatGPT para pesquisas e escrita acadêmicas já foram sentidos no âmbito da pesquisa médica. Em artigo na revista *The Lancet*, indicou-se que há uma “necessidade de se implementar diretrizes robustas para a publicação acadêmica” (Liebrenz, 2023). Tais diretrizes devem envolver questões como “direitos autorais, atribuição, plágio e autoria quando a AI produz o texto acadêmico”. O autor ainda destaca o fato de que textos produzidos por IA podem passar despercebidos pelos softwares de verificação de plágio (Liebrenz, 2023).

No entanto, sabe-se que os modelos atuais de LLM – pelo menos até o momento – não armazenam as respostas dadas a fim de fornecer uma indicação segura de que determinado texto foi realmente produzido por aquela plataforma¹⁶. Os textos produzidos podem diferir de tamanho e complexidade, variando desde uma resposta dada a uma questão até uma simples correção gramatical de um texto produzido por um autor humano. Além dessa variabilidade das respostas, o autor ainda pode utilizar ferramentas diferentes de LLM, treinadas de maneiras distintas, com propósitos variados, o que dificultaria bastante a identificação cruzada entre plataformas.

quando se tornar possível, em lugar de se referirem a elas, referirem-se às decisões judiciais que as aplicaram”. Ver também Losano (2007, pp. 336-337): “As vantagens e as deficiências do *Common Law* são em certa medida simétricas às do direito continental: sua possibilidade de se adaptar ao caso concreto é maior do que nos sistemas jurídicos de normas gerais e abstratas, mas estas são reduzíveis a um corpus limitado e sistemático, enquanto os precedentes tendem a formar uma floresta indestrutível e extensa. A extensão dos precedentes torna às vezes materialmente impossível ter uma visão clara da situação jurídica de certos setores, fazendo vacilar a certeza do direito”.

16 Destacando o fato de que, diante de sua natureza estocástica, ele produz respostas ligeiramente diferentes mesmo diante dos mesmos comandos, além de levar em considerações as interações anteriores do usuário se o comando for dado dentro de uma mesma janela de conversação. Isso torna bastante difícil avaliar com certeza se um texto foi produzido por um LLM, visto que ele também erra quando tenta fazer essa identificação.

Apesar dos resultados altamente satisfatórios no uso desses modelos para tarefas cotidianas, há uma grande dificuldade em entender como realmente eles funcionam¹⁷. Diante da quantidade de dados utilizados para seu treinamento e as numerosas operações realizadas para prever as palavras, é bastante difícil, senão impossível, fazer o caminho reverso de como aquela palavra foi prevista. A quantidade de operações matemáticas realizadas impede que um humano consiga refazer todo o caminho manualmente. Isso afeta algumas atividades de controle que são realizadas. O problema é: como realizar um controle em um algoritmo que consumiu bilhões de informações, que “aprendeu” a produzir texto novo diante de operações matemáticas altamente complexas, diante das próprias limitações da compreensão humana? (Bowman, 2023, p. 6).¹⁸

Tal complexidade, no entanto, pode ser superada com institutos já existentes no ordenamento jurídico brasileiro, como a utilização do princípio da boa-fé objetiva. Os deveres anexos à boa-fé contemplam aqueles de informação, colaboração e proteção, o que ilumina os comportamentos dos mantenedores das plataformas de LLM.

É claro que soluções que impõem mais controle sobre os resultados podem, em tese, diminuir sua eficiência. Seria possível implementar modelos que pudessem armazenar mais informações sobre como o resultado foi produzido – o que envolveria mais camadas de processamento, deixando-o mais lento e, conseqüentemente, mais caro. Há uma escolha ética a ser feita: talvez seja preciso diminuir sua eficiência em prol de mais explicabilidade. Essa explicabilidade pode até

17 Lee e Trott (2023) dão um exemplo que o GPT-4, mesmo não sendo treinado com imagens, conseguiu realizar previsões sobre locais específicos em imagem, organizando respostas visuais sobre como, por exemplo, o local correto de um chifre em um unicórnio. Os autores afirmam, como hipótese, que: *“At the moment, we don’t have any real insight into how LLMs accomplish feats like this. Some people argue that such examples demonstrate that the models are starting to truly understand the meanings of the words in their training set. Others insist that language models are “stochastic parrots” that merely repeat increasingly complex word sequences without truly understanding them”*.

18 Cf. Bowman (2023, p. 6) 6: *“There are hundreds of billions of connections between these artificial neurons, some of which are invoked many times during the processing of a single piece of text, such that any attempt at a precise explanation of an LLMs behavior is doomed to be too complex for any human to understand”*.

mesmo ser necessária para resolver problemas como a análise da possibilidade de terem sido utilizados dados protegidos por direitos autorais para a geração daquela resposta. A questão é que esse novo dever pode ser até mais complexo de se cumprir do que as próprias operações de geração de texto. Em suma, as dificuldades técnicas podem impedir o cumprimento de deveres de explicabilidade e, assim, poderia ser necessário abrir mão deles – dos deveres – em face da própria eficiência dos resultados. A ferramenta acaba por se tornar um oráculo¹⁹. Esse problema envolve a dificuldade de cumprir o dever de informação e explicabilidade nessas plataformas. Essa falta de transparência²⁰ pode impedir que a IA seja utilizada em uma variedade de contextos mais críticos. Basta pensar no dever de fundamentação e observância de precedentes na produção de decisões judiciais²¹ – embora esse não seja o foco da análise aqui realizada. Como utilizar uma ferramenta que pode sugerir decisões se não se sabe exatamente quais os critérios que foram utilizados para sua fundamentação?

Esse dever de transparência também será variado, dependendo da complexidade e uso das ferramentas de LLM, envolvendo tanto informações sobre os resultados, sobre como elas foram treinadas além de sua aderência a certos padrões (inclusive éticos), o que pode envolver, por exemplo, a eliminação de vieses e preconceitos (Diakopoulos, 2020, p. 199). Não há como responsabilizar adequadamente os mantenedores

19 No âmbito da tomada de decisões judiciais, Lummertz (2018) manifestou-se no sentido de que: “A transparência e auditabilidade dos algoritmos e sistemas de inteligência artificial utilizados na tomada de decisões pelo Poder Judiciário e pela Administração Pública evitarão que tais decisões se convertam em verdadeiros vaticínios, como aqueles provenientes do Oráculo de Delfos, manifestações da vontade divina, cujas razões se desconhece, mas que, assim mesmo, deve-se obedecer”.

20 Sobre o tema ver Diakopoulos (2020, p. 198). Segundo o autor: “*Transparency therefore provides the informational substrate for ethical deliberation of a system's behavior by external actors. It is hard to imagine a robust debate around an algorithmic system without providing to relevant stakeholders the information detailing what that system does and how it operates. Yet it's important to emphasize that transparency is not sufficient to ensure algorithmic accountability. Among other contingencies, true accountability depends on actors that have the mandate and authority to act on transparency information in consequential ways*”.

21 No exemplo dado por Bostrom e Yudkowsky (2014, p. 317) os autores apontam alguns valores a serem perseguidos: “*Responsibility, transparency, auditability, incorruptibility, predictability*”.

dessas ferramentas – e tampouco permitir os mais variados exercícios de direitos das pessoas – sem transparência adequada²². Uma possível instrumentalização desse princípio, no âmbito dos direitos autorais, é a disponibilização pelos responsáveis de uma forma de conferir se determinado conteúdo foi ou não utilizado no processo de treinamento da ferramenta. Trata-se de forma relativamente simples de conferência, já que os administradores sabem quais informações foram utilizadas. Por outro lado, a questão se reveste de intensa complexidade quando houver a necessidade de se retirar uma informação do modelo, pois ela implicaria, ao menos em tese, na necessidade de treinar novamente o algoritmo, o que pode demorar semanas ou meses para acontecer²³. Sendo muitas as solicitações de retirada de informações, a tarefa de treinar novamente o algoritmo pode se tornar impossível. Tarefa mais simples seria a de verificar, de antemão, que não se utilizou informações protegidas por direitos autorais, treinando o algoritmo somente com informações não protegidas por direitos autorais, o que, por certo, comprometeria a eficiência da ferramenta.

Em um primeiro momento, aquele que cria uma obra para publicá-la não está em uma relação propriamente obrigacional. Contudo, aquele que publica uma obra, de qualquer natureza, tem, ao menos, o dever geral de não infringir leis ou causar danos a outros autores, violando seus direitos. O autor de obra que se proponha a ser inédita não pode, por isso, realizar contrafação.

22 Ver de maneira geral, a recomendação da Comissão Europeia (2021). *Proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União* [online]. Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF (Accessed: 14 August 2023), p. 13: “O aumento das obrigações de transparência também não afetará desproporcionadamente o direito à proteção da propriedade intelectual (artigo 17º, n.º 2), uma vez que estarão limitadas às informações mínimas necessárias para as pessoas singulares exercerem o seu direito à ação e à transparência necessária perante as autoridades de supervisão e execução, em conformidade com os mandatos destas”.

23 O Parlamento Europeu já se manifestou sobre a necessidade de “desenvolver novos métodos, como, por exemplo, sistemas de aprendizagem adaptáveis que se podem recalibrar após cada entrada”. PARLAMENTO EUROPEU (2020) *Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial*, p. 6.

Destaque-se, ainda que, para ser possível tratar de proteção autoral, há que ser evidenciada originalidade e individualidade, mesmo que em nível mínimo, inclusive, alcançando direitos de autor daquele que programa, bem como em seus *inputs*, segundo leciona Marilù Capparelli:

Para que possa ser dada proteção no sentido da normativa de direito de autor, a obra deve ser dotada de caráter criativo, de um nível mínimo de originalidade e individualidade que reflita a personalidade e o engenho de seu autor, e que distingue a obra em si da obra que lhe precedeu. [...] A originalidade do ato criativo poderá, agora, ser reenviada para o momento antecedente da obra, ou seja, no momento da programação da máquina, coincidente com a decisão e organização da criação da obra: seria, assim, nas escolhas completas pelo programador e nos *inputs* feitos para a base da programação da IA, que se poderia reenviar aquele momento de criatividade necessário para atribuir o direito de monopólio exclusivo no que toca à obra de criatividade, em tal medida reconduzível ao intelecto do programador; o qual poderá, portanto, ser considerado o titular dos direitos de autor sobre a obra (traduziu-se) (Capparelli, 2020, pp. 335-342).

Nesta linha, que o Copyright, Design and Patent Act de 1988, estabelece que é autor aquele que tomou as medidas necessárias para o processo criativo da obra gerada, em disposição de vanguarda: “No caso de uma obra literária, dramática, musical ou artística gerada por computador, o autor deve ser considerado a pessoa que realizou os arranjos necessários para a criação da obra” (Reino Unido, 1988).

É inegável que a criação intelectual, quando feita com o apoio de uma inteligência artificial, desafia a própria natureza do ato de criar. Pontes de Miranda (1956, p. 13) já dizia que “a inserção do elemento psíquico próprio é que faz a criação intelectual”. Há várias formas de analisar a questão da propriedade intelectual e a IA. A primeira é a proteção dos algoritmos que instrumentalizam os LLMs, o que não parece trazer tantos problemas, eis que se trata de software, contando com regras de proteção específicas. Porém, quando se fala dos dados utilizados para o treinamento dos algoritmos, tem-se uma complexa teia de situações jurídicas envolvidas. O primeiro

diz respeito ao fato de que tais modelos usam dados que estão disponíveis na Internet. Ocorre que, diante da opacidade das técnicas utilizadas, não se sabe ao certo quem são os titulares de tais dados, visto que se pode estar diante de dados protegidos por direitos autorais, dados pessoais protegidos por leis específicas e, até mesmo, imagens e dados pessoais, amparadas pelos direitos da personalidade e a Lei Geral de Proteção de Dados. Essa miríade de situações, diante da falta de clareza e transparência por parte dos administradores de tais plataformas, impossibilita que os titulares de direitos saibam que seus dados estão sendo tratados e se oponham às atividades que eventualmente violem a Lei.

Não se perca de vista, também, que a própria atividade de treinamento poderia ser analisada sob a ótica do Direito das Obrigações, no âmbito do enriquecimento sem causa²⁴. Mesmo que as obras utilizadas para o treinamento não possam ser identificadas nos *outputs*, os mantenedores das ferramentas podem utilizar obras protegidas por direitos autorais para a criação dos algoritmos, o que lhes trará evidentes benefícios financeiros. Nesse sentido, antes mesmo das discussões atinentes aos direitos autorais, há uma discussão que deve ser feita sobre a possibilidade de aplicação do instituto do enriquecimento sem causa ao caso.

O ponto de reflexão passa a ser se a utilização de criações intelectuais para treinar o algoritmo exige que o seu mantenedor respeite os direitos autorais aplicáveis ou se essa atividade isenta o dono do algoritmo de respeitar os direitos dos autores. Parece claro que o uso de obras intelectuais neste contexto não pode ser realizado sem o respeito dos direitos autorais. Esta hipótese violaria até um paradigma ético do uso das IA que deveria respeitar, em última análise, a própria ideia de colaboração (Benanti, 2018, p. 55)²⁵ e o bem comum. Além

24 Cf. o art. 884 do CC: “Aquele que, sem justa causa, se enriquecer à custa de outrem, será obrigado a restituir o indevidamente auferido, feita a atualização dos valores monetários”.

25 Cf. Benanti (2018, p. 55): “*In altre parole, le intelligenze artificiali (machinae sapientes) non sono degli avversari evolutivi dell'homo sapiens bensì strumenti (artefatti) che devono essere pensati come cooperativi alla persona. Di fatto le intelligenze artificiali devono essere realizzate per aumentare la capacità cognitiva, che è prerogativa unica e peculiare dell'uomo, e non sostituirsi mai a questi*”.

disso, parece até mesmo haver um abuso de direito dos responsáveis pelos algoritmos que treinam os algoritmos de forma a não permitir que os autores das obras utilizadas sequer saibam que suas obras foram utilizadas.

Pense-se em um exemplo mais claro. Sabe-se que o Tratado de Direito Privado de Pontes de Miranda está disponível para download na Internet. Diante disso, seria lícita a atuação de treinar uma inteligência artificial com todos os livros do autor e disponibilizar, assim, um chat para responder perguntas com base em tais obras? E se assim fizesse, a Lei brasileira de direitos autorais seria aplicável? Se é claro que a Lei deve ser respeitada neste contexto, o mesmo não poderia ser dito acerca de todas as situações em que se utilizam obras de outros autores para treinar algoritmos de IA?

Assim, compreende-se que a observação aos direitos de autor passa estar entre as diretrizes a serem respeitadas pelos mantenedores dos LLMs. A Lei dos Direitos Autorais, em seu artigo 7º, manifesta a sua vocação futurista, na medida em que estabeleceu que são protegidas as criações de espírito, “por qualquer meio ou fixadas em qualquer suporte”, “conhecido ou que se invente no futuro”²⁶, e, portanto, voltando-se ao porvir, com incidência plena, imediata e inarredável sobre o campo das novas tecnologias e, no caso, sobre os LLMs.

Devem ser observados os direitos autorais e conexos, tanto no que tange ao âmbito moral, como patrimonial, sobretudo os artigos 24 a 29 da Lei dos Direitos Autorais.

Outra importante diretriz se apresenta na medida da utilização para alimentação da IA de obras que já estão em domínio público, ou seja, como no caso de já ter transcorrido mais de setenta anos da morte do criador, a contar do primeiro dia do ano seguinte ao falecimento, nos termos do artigo 41, da Lei de Direitos Autorais. Incluem-se aí também as exceções presentes no art. 46 da referida lei ou, ainda, o uso de obras que estão em licenciadas no modelo de *Creative Commons*, ade-

26 Compartilhando sobre Direitos Autorais, voltadas às plataformas em Colombo, C. (2020) ‘Direitos Autorais na Indústria Criativa e conteúdo ilícito gerado por terceiros’, *Migalhas*, 30 September [online]. Available at: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/334150/direitos-autorais-na-industria-criativa-e-conteudo-ilicito-gerado-por-terceiros> (Accessed: 13 August 2023).

quando-se às suas licenças, se para fins lucrativos ou não, e estando vedada a utilização de licenças que não permitam obras derivadas, já que o *output* poderá ser identificado como tal.²⁷ A legislação autoral admite, de forma expressa, ainda, que seja possível a utilização de textos para paródias, na linha do artigo 47 da LDA, quando as “reproduções de obras originárias não impliquem em descrédito ao autor”.

Porém, acerca de licenças específicas de *Creative Commons* ou de código fonte aberto (*open source*), é preciso também observar as próprias limitações de uso dessas licenças. É comum que nestes casos haja a necessidade de que as obras derivadas sejam disponibilizadas também com as mesmas licenças. Um caso recente envolveu o treinamento da ferramenta Copilot, da Microsoft, que ajuda programadores a criarem código com o uso da IA. A solução funciona como um ChatGPT específico para a criação de códigos fonte de softwares. Ocorre que a solução foi treinada com códigos *open source* armazenados em um repositório, também da Microsoft, chamado de GitHub. A comunidade de desenvolvedores de software se insurgiu contra o fato de que quando os desenvolvedores publicaram seu código no formato *open source*, o uso para treinamento de IA feriria os termos da própria licença, já que a solução é um projeto comercial (o que é vedado pela licença), além de desrespeitar os

princípios éticos de compartilhamento que levaram os desenvolvedores a distribuir seu código abertamente em primeiro lugar. Em segundo lugar, houve críticas ao código que é sugerido pelo Copilot, com alguns programadores alegando que ele reproduz o próprio código deles de volta (Guedamuz, 2023).

Isso demonstra a dificuldade de compatibilizar as soluções atuais de proteção aos direitos autorais com as novas possibilidades trazidas pelas ferramentas de LLM, tanto no processo de treinamento quanto na produção de *outputs*. Diante do caso apontado, portanto, há que se respeitar os limites

27 CREATIVE COMMONS [online]. Available at: https://creativecommons.org/licenses/?lang=pt_BR (Accessed: 13 August. 2023).

das licenças abertas, inclusive quanto a possíveis limitações de seu uso comercial.

3.2. Novos deveres Éticos-Jurídicos para sua utilização

Diante do exposto, é preciso avaliar quais são os novos deveres de autores diante do uso da IA para a produção acadêmica. O primeiro ponto envolve a análise de como o conceito de autoria varia com essas novas tecnologias. Note-se que a situação analisada envolve o uso de LLMs por autores humanos que o utilizam para apoiar ou complementar textos autorais. Em recente artigo intitulado *ChatGPT listed as author on research papers*, a comunidade acadêmica analisou o papel do ChatGPT, que havia sido indicado em pelo menos quatro publicações, como autor, trazendo questionamentos sobre “o futuro dos ensaios universitários” e a “produção de pesquisa”. O Time da Nature questionou editores sobre o ChatGPT, sendo que se manifestaram que o LLM “não preenche os critérios de estudo autoral, porque não tem responsabilidade pelo conteúdo e integridade do trabalho científico”, e, que “[...] somente pessoas podem ser listadas” (Stokel-Walker, 2022). Os editores-chefe da Science e da Nature referiram que “LLMs não atendem padrão de autoria” (Stokel-Walker, 2022). E, no caso de ser utilizado, o LLM deverá ser indicado na seção de métodos, uma vez que texto feito por IA, sem a citação adequada, pode ser compreendida como plágio (Stokel-Walker, 2022).²⁸

Importante preocupação envolve formas de evitar plágios nos trabalhos acadêmicos. Concorde-se com a afirmação de que “O plágio trata-se de uma questão ética, antes do que jurídica” (Pithan and Vidal, 2013, p. 78). No entanto, a ideia de plágio como reprodução de passagens de texto de outros autores sem a devida citação, ou reprodução de conceitos, ideias ou teorias sem a indicação da fonte, parece ser uma das vias

28 Segundo as regras de publicação de Nature: “Large Language Models (LLMs), such as ChatGPT, do not currently satisfy our authorship criteria. Notably an attribution of authorship carries with it accountability for the work, which cannot be effectively applied to LLMs. Use of an LLM should be properly documented in the Methods section (and if a Methods section is not available, in a suitable alternative part) of the manuscript” (Stokel-Walker, 2022).

a serem consideradas diante da IA. Talvez uma aproximação mais apropriada seria considerar os LLM como ferramentas de apoio com diferentes níveis de intervenção no texto. A depender do seu nível de interação, elas podem atuar de forma semelhante a um *ghost writer* (Zanini, 2015)²⁹. Isso porque o conteúdo gerado não existia propriamente até sua geração (embora as condições para sua geração já estejam dadas no algoritmo treinado), a não ser que a IA realmente reproduza um texto já criado anteriormente, o que é sim possível a depender da instrução dada pelo usuário. Assim como ocorre com o trabalho de um *ghost writer*, há uma colaboração entre ele e um terceiro que fornece informações, sendo que aquele as organiza em formato literário³⁰.

Os níveis de interação de humanos com as ferramentas de IA podem ser muito variados, o que envolve intervenções mais ou menos intensas no texto produzido. Há um escalonamento na interferência que a IA pode realizar na produção de um texto. Graus mais intensos de interferência levam a uma afetação na atribuição de autoria do texto. Uma situação básica, que não parece envolver grandes problemas jurídicos (ou éticos) de atribuição de autoria ou até mesmo de plágio, ocorre quando o autor apenas faz uma correção gramatical de

29 Ver Zanini (2015): O problema do *ghost writer* nos direitos autorais demanda também reflexões profundas. Zanini afirma que se trata de uma questão que envolve a análise do “problema da renúncia aos direitos da personalidade do autor” (Zanini, 2015, p. 293). O autor ainda aponta que na doutrina alemã o problema é tratado no âmbito do “não exercício obrigacional e limitado no tempo do direito de paternidade” (Zanini, 2015, p. 295).

30 Lembrando do famoso caso brasileiro do livro escrito por um *ghost writer* e pela ex prostituta Bruna Surfistinha. A história chegou a ser transformada em filme. Para mais informações, inclusive no âmbito jurídico, ver Virtuoso, B. and Graton, N. A. (2015) ‘Ghost-Writer: uma breve análise a partir do caso Bruna Surfistinha’. *GEDAI*, 30 May [online]. Available at: <https://www.gedai.com.br/ghost-writer-uma-breve-analise-a-partir-do-caso-bruna-surfistinha/> (Accessed: 10 August 2023). O caso chegou até o STJ (REsp 1.387.242), conforme resumo de Virtuoso e Graton: “o recorrente apenas redigiu o texto, fazendo as correções necessárias, de forma que **não traduz o desenvolvimento literário que configura a autoria**. Além disso, o escritor tinha ciência que havia sido contratado pela editora para prestar serviços como Ghost-writer, profissional especializado em redação de textos para terceiros abdicando da paternidade da obra”. Também citando o caso Zanini, Leonardo Estevam de Assis. *Ibidem*, p. 299, apontando que: “Não temos dúvida que coube ao autor a criação do livro, pois foi ele quem selecionou as situações que reputou mais interessantes, criou uma narrativa que intercalou fatos ocorridos antes e depois da fuga de Raquel de sua casa, bem como imprimiu na obra a sua personalidade, a sua forma de ver a vida das personagens abordadas”.

um texto próprio. Nesta situação, o autor submete um texto totalmente original para avaliação gramatical. Com isso, a ferramenta atuará como um corretor, semelhante ao que acontece com a correção automática nas ferramentas tradicionais de edição de textos. Situações semelhantes ocorrem quando o autor busca a substituição de palavras por sinônimos ou até mesmo quando utiliza a ferramenta para a tradução de textos.

O nível de interferência começa a se aprofundar, demandando mais reflexão, quando é solicitado à ferramenta a melhoria de um texto autoral, eis que pode ser alterada a sua estrutura. Tome-se como exemplo o parágrafo anterior. Utilizando-se o GPT-4, forneceu-se a seguinte instrução: “Melhore e resuma o parágrafo a seguir”, obtendo-se como resposta:

A interação humana com ferramentas de IA varia em intensidade, afetando a atribuição de autoria do texto. Em situações básicas, onde o autor usa a IA para correções gramaticais ou busca de sinônimos em um texto original, poucas questões jurídicas ou éticas surgem. Isso se assemelha à correção automática de editores de texto tradicionais ou ao uso de tradutores automáticos.

Note-se que o sentido do parágrafo não foi alterado, as palavras principais foram mantidas e o resumo manteve a essência do texto. Esse tipo de interação poderia ser entendido como uma citação indireta³¹, se realizado por outra pessoa que não o autor – demandando sempre a citação do texto original. No entanto, como a ideia de citação indireta envolve um texto baseado na obra de outro autor, não haveria a necessidade do próprio autor afirmar que fez um resumo de um texto seu.

Acredita-se que não há questões controversas para o próprio autor quando ele solicita a melhoria ou o resumo de textos próprios. Ocorre situação semelhante quando um autor submete sua dissertação ou tese para um revisor profissional. Neste caso, não há questões éticas envolvendo problemas de confusões de autoria³² ou a possibilidade de atribuição de

31 Lembrando que a citação direta é a “transcrição textual de parte da obra do autor consultado” e a citação indireta é o “texto baseado na obra do autor consultado”, cf. ABNT NBR 10520 (ABNT, 2013).

32 Lembrando que o art. 11 da lei 9.610/98 define o autor como “pessoa física criadora de obra literária, artística ou científica”.

co-autoria. Denominamos esses usos da IA como *usos de apoio primário*.

Os termos de uso da OpenAI concedem ao usuário a titularidade dos *outputs*, permitindo que sejam utilizados inclusive para fins comerciais, o que envolve inclusive a publicação³³. Considerando que a OpenAI oferece uma ferramenta de geração de textos e suas políticas – com força obrigacional, mesmo que o vínculo se dê por adesão – transferem todos os direitos de uso do texto produzido, não parece ser difícil entender que a ferramenta funciona como um *ghost writer*. Veja-se aqui até a dificuldade de se analisar possíveis renúncias de direitos do autor – eis que a ferramenta não é humana –, visto que, ao menos pela Lei atual, não haveria como se atribuir autoria a uma solução de inteligência artificial. É claro que esta solução jurídica não afasta o aspecto ético do autor mencionar o uso de ferramentas de IA quando houver um grau intenso de intervenção no texto produzido. E a solução é diferente daquela em que alguém usa um buscador para responder a uma questão ou para buscar uma informação qualquer. A diferença principal é que o buscador entrega, na maioria das vezes, as fontes originais que embasaram suas respostas, permitindo ao autor citar diretamente a informação encontrada por meio do buscador. Já com as ferramentas de IA, na maioria das vezes não se obtém a fonte utilizada para a resposta dada por ela e nem poderia, diante das características de geração de conteúdo, que estão mais relacionadas a regras estatísticas que organizam o texto baseados na probabilidade da próxima palavra naquele contexto do que o apontamento de uma fonte específica. É claro que devem ser levadas em consideração as fontes usadas para treinar o algoritmo, porém o número de

33 OpenAI (2023). OpenAI, Terms of Use, 14 March [online]. Available at: <https://openai.com/policies/terms-of-use>. (Accessed: 2 August 2023): “Your Content. You may provide input to the Services (“Input”), and receive output generated and returned by the Services based on the Input (“Output”). Input and Output are collectively “Content.” As between the parties and to the extent permitted by applicable law, you own all Input. Subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output. This means you can use Content for any purpose, including commercial purposes such as sale or publication, if you comply with these Terms. OpenAI may use Content to provide and maintain the Services, comply with applicable law, and enforce our policies. You are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms”.

fontes usadas torna rarefeita sua influência no conteúdo gerado, sendo bastante difícil, em alguns contextos, que a ferramenta aponte quais fontes foram utilizadas para gerar o resultado (a não ser que se peça expressamente por elas).

As fontes de treinamento servem também, não se esqueça, para organizar os vetores de palavras em contextos e aproximações, não somente para trazer informações. Isso faz com que as informações usadas no processo possam não aparecer explicitamente no resultado, servindo como fonte para a organização dos vetores e, em última análise, de contextos. Apreende-se também a estrutura gramatical do texto sem, necessariamente, refletir, nos resultados, o seu conteúdo. São dimensões diferentes na geração de resultados: a coerência do texto versus seu amparo em fontes científicas ou factuais.

Não se descuide que há também uma criatividade do autor nos *inputs* que ele dá, o que aproxima ainda mais a situação com a atuação de um *ghost writer*. O usuário pode ir informando situações e pedindo, pouco a pouco, que a ferramenta vá escrevendo uma história com base naquelas informações. Pode-se dar informações sobre personagens, características, fatos, relacionamentos entre personagens, etc. Porém, a criatividade do autor – ou sua falta – pode fazer com que ele peça para a ferramenta imitar estilos³⁴. Níveis de interação mais aprofundados da ferramenta ocorrem quando o usuário humano solicita a criação de novos textos somente com a introdução de apenas uma instrução. Nestes casos é altamente recomendável que a ferramenta de IA seja citada explicitamente

34 CHATGPT-4. Prompt: “Escreva um pequeno poema inédito copiando o estilo do poeta Mário Quintana”. Geração em: 4 ago. 2034. <https://chat.openai.com/>. Resposta: Na calçada da memória, encontro-me sozinho, / Onde a lâmpada do tempo tece o crepúsculo do dia. / Ora distante, ora perto, tal qual um passarinho, / Que entre árvores esvoaça, onde sua canção porfia. Sob o manto de estrelas, meu pensamento divaga, / Enquanto o silêncio da noite a tudo permeia. / No espelho d'água, o luar em seu reflexo se embriaga, / Em doce melancolia, na qual a alma se enleia. Ah, as ruas desta vida, trilhas do meu viver, / São repletas de risos, de lágrimas e poesia. / Que se desenrola, como fita ao vento, a correr, / Na tela infinita do tempo, onde o sonho desafia. Mas na roda da existência, onde tudo se desdobra, / Nem sempre se alcança a linha da eternidade. / O importante, meu amigo, não é o que se cobra, / Mas o amor com que se vive, em toda a sua amplitude.

como meio na obtenção daquele resultado, sendo as informações geradas indicadas como citação direta ou indireta³⁵.

Questões mais complexas emergem quando o usuário solicita informações factuais e conceituais³⁶. A depender dos dados que treinaram a ferramenta, os resultados podem ser mais ou menos fidedignos, ocorrendo até o apontamento de informações incorretas³⁷. Por outro lado, existem modelos que são treinados para atividades mais específicas, como modelos utilizados para a medicina. Acerca dos *layers* que se indicou anteriormente, eles possuem funções de ajustes de contextos. Alguns problemas aparecem quando a ferramenta produz *outputs* que fazem sentido gramaticalmente, mas que não possuem respaldo na realidade fática ou científica. A própria OpenAI, desenvolvedora do modelo GPT, reconhece tais limitações e ainda aponta a necessidade de um cuidado na análise dos resultados

[...] em contextos de alto risco, com o protocolo exato (como revisão humana, fundamentação com contexto adicional ou evitando usos de alto risco completamente) correspondendo às necessidades de aplicações específicas³⁸.

Um dos testes feitos pelos autores deste artigo consistiu no solicitação de indicação de músicas que tivessem como tema a educação. Entre os resultados, foram trazidas duas músicas que não existiam, fenômeno chamado de “alucinação”. Diante desse risco de respostas que não possuem base na realidade fática ou científica, modelos mais aprimorados poderiam inserir outros *layers* de apuração ou confirmação, podendo até utilizar a própria internet para tais conferências. Não se trata somente de utilizar *plugins* para que a ferramenta busque informações na Internet, mas que ela tenha a capacidade de corrigir suas respostas, como um último nível de “decantação” do conteúdo, visando eliminar incorreções ou inconsistências

35 Como se fez no presente texto.

36 Como, por exemplo, quando morreu determinada personalidade, quem foi o presidente do Brasil em tal ano ou quem é o autor de tal teoria.

37 Já que a ferramenta pode meramente privilegiar a coerência gramatical do texto, desprezando a correção informacional ou factual das respostas apresentadas.

38 OPENAI (2023) *GPT-4 Technical Report*, p. 10 [online]. Available at: <https://arxiv.org/pdf/2303.08774.pdf> (Accessed: 3 August 2023).

fáticas. Trata-se de incluir uma outra dimensão no algoritmo que avalie, além do sentido gramatical, sua correção factual e científica. Note-se que isso adiciona uma camada bem mais complexa de conferência, já que o modelo precisaria pesquisar e avaliar todos os *outputs* dados, o que deixaria a entrega das respostas mais demorada³⁹. Isso já ocorre, de certa forma, com as ferramentas de moderação. Simples interações com o ChatGPT solicitando temas mais controversos, ou até mesmo o uso de palavras⁴⁰, trazem *outputs* que refletem decisões éticas dos desenvolvedores em limitar o alcance do conteúdo.

Não se perca de vista, também, a hipótese aventada por Marcelo Bürger e Rafael Corrêa (2020, p. 531), quando invocam o art. 40 da Lei 9.610/98 ao prever a possibilidade de proteção de obras anônimas ou pseudo anônimas. Os autores apontam que os direitos patrimoniais de tais obras serão gozados a quem as publicar. No entanto, essa hipótese não se encaixa completamente no caso, eis que o ser humano quando interage com as ferramentas de IA está a lhes dar direção, sendo mais adequada, em nosso sentir, a hipótese de proteção semelhante à aplicada às situações de participação de *ghost writer*, sempre observando o grau de interação do ser humano, conforme se expôs.

Nesse sentido, devem os LLMs, observar os seguintes deveres éticos-jurídicos:

1) observar os direitos de autor, no âmbito moral, com a obrigação de informar quais os textos foram utilizados, no processo de treinamento, como *input* na Inteligência Artificial, nos termos do artigo 24, da Lei de Direitos Autorais;

39 Nada impede que a ferramenta adicione uma opção de conferência, para que o usuário somente a acione quando achar necessário. Essa opção não seria ativada, por exemplo, quando se busca a mera correção de um texto, eis que utilizaria somente as funcionalidades gramaticais da solução.

40 CHATGPT-4. Prompt: Indique quais os palavras mais ofensivas para se xingar alguém. Acesso em: <https://chat.openai.com/>. Acesso em 13 ago de 2023. A resposta foi: “Desculpe, mas eu não vou fornecer uma lista de palavras ou incentivar o uso de linguagem ofensiva. Se você tiver dúvidas ou necessidades relacionadas a linguagem, educação, cultura ou outros tópicos, estou aqui para ajudar!”. Note-se que apesar do modelo possuir a capacidade de produzir textos envolvendo palavras, a moderação da plataforma gerou uma contenção no discurso.

2) caso seja utilizado obra total ou parcial, que gerará obra derivada, a autorização dos criadores das obras primíge-nas, nos termos do artigo 29, da Lei dos Direitos Autorais;

3) caso sejam utilizados pequenos trechos, a LDA não exige autorização prévia e expressa, mas, em todos os casos, o *ou-tput* deve citar quem apresentou a definição, os exemplos ou ou-tras informações, para que não haja violação a direito de autor. É o que pode ocorrer, na maioria das vezes, quando a resposta se utiliza de múltiplos textos, devendo apenas referenciá-los;

4) a utilização de textos em domínio público, licenciados por *creative commons* ou cobertos pelas exceções do art. 46 da Lei de Direitos Autorais, como *inputs*, poderá minimizar violações a direitos patrimoniais de autor, sempre sendo preservado o di-reito autoral. Neste caso, deve-se observar também as próprias licenças *creative commons* quando limitam seu uso comercial, o que se aplica também às licenças de código aberto.

5) implementar controles que não reproduzam respostas exatamente iguais a materiais protegidos por direitos autorais sem a indicação da fonte. Isso eliminaria (ou ao menos diminui-ria, do ponto de vista da plataforma) a presença de plágios, per-mitindo que a ferramenta possa ser utilizada de forma mais se-gura pelos usuários. Não se pode perder de vista que o aumento dos controles pode afetar sensivelmente a eficiência e o custo de manutenção da plataforma, o que importa em um problema econômico. Há uma tendência da ferramenta se tornar mais cara e complexa sempre que implementar controles para cum-prir as Leis. Em última instância, isso diminuiria os lucros dos seus mantenedores, daí o problema econômico a ser enfrenta-do. Por outro lado, é preciso promover um ambiente institucio-nal que permita o desenvolvimento das ferramentas de inteli-gência artificial e de suas potencialidades com a necessidade “de assegurar um elevado nível de proteção de DPI (Direitos de Propriedade Intelectual)” (Parlamento Europeu, 2020).

Dessa forma, há a necessidade de se estabelecer uma supervisão pela curadoria de dados (Bender, Gebru, McMillan-Major and Shmitchell, 2021). Ora, se a Inteligência Artificial é como uma criança, “aprende com a experiência” (Ruffolo, 2021), dar a ela o dado hígido e curado, identificando-o e clas-

sificando-o, para a realização de uma determinada tarefa, mitigará violações aos direitos da pessoa humana.

4. Considerações finais

Com a chegada dos LLMs, é importante que se promova uma releitura dos Direitos de Autor, para que siga sendo promovida a tutela destes vulneráveis. No âmbito moral, deve ser observada a obrigação de fornecer os textos utilizados como *input* no processo de treinamento, na Inteligência Artificial, para atendimento do artigo 24, da Lei de Direitos Autorais. Na hipótese da utilização de obra total ou parcial, que gerará obra derivada, deverá ser coletada a autorização dos criadores das obras primárias, nos termos do artigo 29, da Lei dos Direitos Autorais. Em sendo utilizado pequeno trecho, o *output* deve citar quem apresentou a definição, os exemplos, enfim, para que não haja violação a direito de autor, não sendo exigida, no caso, autorização prévia e expressa, eis que hipótese do artigo 46, da LDA. Dessa forma, critério bastante importante é a identificação de elementos substanciais de uma obra no resultado (*output*), que, a depender da proporcionalidade, deve ser determinante para a coleta da autorização prévia e expressa ou, como acima referido, se pequeno trecho for, deve ser caso de citação, ainda que indireta. A utilização de textos em domínio público ou *creative commons*, como *input*, no processo de treinamento, poderá minimizar violações a direitos patrimoniais de autor, sempre sendo preservado o direito autoral. O que se espera é que editores de revistas científicas criem seus guias de conformidade para utilização dos LLMs, preservando a autonomia humana e estabelecendo os limites destas ferramentas.

Referências

Associação Brasileira de Normas Técnicas (2013) *ABNT NBR 10520: Informação e documentação – Citações em documentos – Apresentação*. Rio de Janeiro: ABNT.

- Benanti, P. (2018) *Le macchine sapienti: Intelligenze artificiali e decisioni umane*. Bologna: Marietti.
- Bender, E. M., Gebru, T., Mcmillan-Major, A. and Shmitchell, S. (2021) 'On the dangers of stochastic parrots Can language models be too big?' *AccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, March, pp. 610-623 [online]. Available at: <https://dl.acm.org/doi/10.1145/3442188.3445922>. (Accessed: 12 August 2023).
- Bonadio, E., Dinev, P. and McDonagh, L. (2022) 'Can artificial intelligence infringe copyright? Some reflections' in Aboott, R. *Research Handbook on Intellectual Property and Artificial Intelligence*. Cheltenham: Edward Elgar.
- Bostrom, N. and Yudkowsky, E. (2014) 'The ethics of artificial intelligence' in Frankish, K. and Ramsey, W. M. *The Cambridge Handbook of Artificial Intelligence*. Cambridge: Cambridge University Press.
- Bowman, S. R. (2023) *Eight Things to Know about Large Language Models* [online]. Available at: <https://cims.nyu.edu/~sbowman/eightthings.pdf>. (Accessed: 13 August 2023).
- Bürger, M. L. F. de M. and Corrêa, R. (2020) 'Dos pincéis aos algoritmos: a titularidade das expressões artísticas e criativas resultantes da aplicação da inteligência artificial' in Erhardt Júnior, M., Catalan, M. and Malheiros, P. (coords.) *Direito Civil e Tecnologia*. Belo Horizonte: Fórum.
- Capparelli, M. (2020) 'Le nuove frontiere del diritto d'autore alla prova dell'Intelligenza Artificiale' in Ruffolo, U. (coord) *Intelligenza Artificiale: il diritto, i diritti, l'etica*. Milão: Giuffrè Francis Lefebvre.
- Carvalho, A. G. P. (2021) 'Defesa da concorrência e inteligência artificial: Desafios impostos pela colusão motivada por algoritmos de fixação de preços' in Mendes, L. S., Alves Jr, S. and Doneda, D. (coords.) *Internet & Regulação*. São Paulo: Saraiva.

- Colombo, C. (2020) 'Direitos Autorais na Indústria Criativa e conteúdo ilícito gerado por terceiros'. *Migalhas*, Migalhas de Responsabilidade Civil, 30 September [online]. Available at: <https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/334150/direitos-auto-rais-na-industria-criativa-e-conteudo-ilicito-gerado-por-terceiros> (Accessed: 13 August 2023).
- Comissão Europeia (2021). *Proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de inteligência artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União* [online]. Available at: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF (Accessed: 14 August 2023).
- David, R. (2002) *Os grandes sistemas do direito contemporâneo*. São Paulo: Martins Fontes.
- Diakopoulos, N. (2020) 'Accountability, Transparency, and Algorithms' in Dubber, M. D., Pasquale, F. and Das, S. (eds.). *The Oxford Handbook of Ethics of AI*. Oxford: Oxford University Press.
- Floridi, L. (2022) *Etica dell'intelligenza artificiale*. Milano: Raffaello Cortina.
- Grimmelmann, J. (2016) 'Copyright for Literate Robots'. *Iowa Law Review*, 101, pp. 657-681.
- Guadamuz, A. (2023) 'A Scanner Darkly: Copyright Liability and Exceptions in Artificial Intelligence Inputs and Outputs'. *SSRN*, 26 February, Draft version 1.5. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4371204 (Accessed: 13 August 2023).
- Lee, T. B. and Trott, S. (2023) 'A jargon-free explanation of how AI large language models work'. *Ars Technica*, 31 July [online]. Available at: <https://arstechnica.com/science/2023/07/a-jargon-free-explanation-of-how-ai-large-language-models-work> (Accessed: 13 August 2023).

- Liebreznz, M. (2023) 'Generating scholarly content with ChatGPT: ethical challenges for medical publishing'. *The Lancet Digital Health*, 6 February. Available at: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(23\)00019-5/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00019-5/fulltext) (Accessed: 13 August 2023).
- Losano, M. G. (2007) *Os grandes sistemas jurídicos: Introdução aos sistemas jurídicos europeus e extra-europeus*. São Paulo: Martins Fontes.
- Lummertz, H. (2018) 'Algoritmos, inteligência artificial e o Oráculo de Delfos'. *Jota, Tecnologia*, 12 October [online]. Available at: <https://www.jota.info/opiniao-e-analise/artigos/algoritmos-inteligencia-artificial-e-o-oraculo-de-delfos-12102018>. (Accessed: 12 August 2023).
- MIT (2021) 'A corrida para entender o mundo emocionante e perigoso dos modelos de linguagem para inteligência artificial'. *MIT Technology Review*, Inteligência Artificial, 4 June. Available at: <https://mittechreview.com.br/a-corrida-para-entender-o-mundo-emocionante-e-perigoso-dos-modelos-de-linguagem-para-inteligencia-artificial/> (Accessed: 13 august 2023).
- MIT (2023a) 'ChatGPT faz com que as pessoas escrevam melhor'. *MIT Technology Review*, Humanos e Tecnologia, 11 August [online]. Available at: <https://mittechreview.com.br/chatgpt-faz-com-que-as-pessoas-escrevam-melhor>. (Accessed: 11 august 2023).
- MIT (2023b) 'Como o texto gerado por Inteligência Artificial está envenenando a Internet', *MIT Technology Review*, Inteligência Artificial, 6 January [online]. Available at: <https://mittechreview.com.br/como-o-texto-gerado-por-inteligencia-artificial-esta-envenenando-a-internet/> (Accessed: 11 August 2023).
- OPENAI (2023) *GPT-4 Technical Report*, 27 March [online]. Available at: <https://arxiv.org/pdf/2303.08774.pdf> (Accessed: 11 august 2023).

- PARLAMENTO EUROPEU (2020) *Resolução do Parlamento Europeu, de 20 de outubro de 2020, sobre os direitos de propriedade intelectual para o desenvolvimento de tecnologias ligadas à inteligência artificial* [online]. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_PT.pdf (Accessed: 11 august 2023).
- Pithan, L. H. and Vidal, T. R. A. (2013) 'O plágio acadêmico como um problema jurídico e pedagógico'. *Direito & Justiça*, 39(1), pp. 77-82, jan./jun.
- Pontes de Miranda, F. C. (1956) *Tratado de Direito Privado*. 2nd edn. Rio de Janeiro: Borsoi, T. XVI.
- Putini, J. (2010) 'Papagaios são isolados em zoológico após ofenderem vários visitantes', *R7*, Hora 7, [online]. Available at: <https://noticias.r7.com/hora-7/fotos/papagaios-sao-isolados-em-zoologico-apos-ofenderem-varios-visitantes-02102020> (Accessed: 12 August 2023).
- Reale, M. (1994) *Lições Preliminares de Direito*. São Paulo: Saraiva.
- Reino Unido (1988) *Copyright, Designs and Patents Act 1988* [online]. Available at <https://www.legislation.gov.uk/ukpga/1988/48/contents> (Accessed: 11 September 2023).
- Ruffolo, M. (2022) 'The Role of Ethical AI in Fostering Harmonic Innovations that Support a Human-Centric Digital Transformation of Economy and Society' in Cicione, F., Filice, L. and Marino, D. (eds.) *Harmonic Innovation. Lecture Notes in Networks and Systems*, 282. Cham: Springer. https://doi.org/10.1007/978-3-030-81190-7_15.
- Stokel-Walker, C. (2022) 'ChatGPT listed as author on research papers'. *Nature*, Berlim, 613, 26 January [online]. Available at: <https://www.nature.com/articles/d41586-023-00107-z> (Accessed: 11 August 2023).
- Tribunal de Contas da União (2020) *Manual de Direitos Autorais*. Brasília. Available at: https://portal.tcu.gov.br/data/files/57/72/86/60/35FA6710FE28B867E18818A8/Manual%20Direitos%20Autorais%202020_Web.pdf. (Accessed: 13 August 2023).

Virtuoso, B. and Graton, N. A. (2015) Ghost-Writer: uma breve análise a partir do caso Bruna Surfistinha. *GEDAI*, 30 May [online]. Available at: <https://www.gedai.com.br/ghost-writer-uma-breve-analise-a-partir-do-caso-bruna-surfistinha/>. (Accessed: 10 august 2023).

Walsh, T. (2022) *Machines behaving badly: The morality of AI*. Cheltenham: Flint, versão Kindle.

Zanini, L. E. A. (2015) *Direito de Autor*. São Paulo: Saraiva.

VIOLENCIAS EN LAS REDES SOCIALES: LA COMPETENCIA DESIGUAL POR SU SENTIDO

Nelson Arteaga Botello¹
DOI: 10.29327/5312711.1-6

Resumen: Se analiza cómo la mayor parte de las investigaciones que abordan la relación entre violencia y redes sociales comparten una perspectiva utilitarista que atribuye a estas últimas una racionalidad que al mismo tiempo despersonalizada y cohesiona horizontalmente. Por un lado, expresan la racionalización de la violencia, el acoso y la dominación; por el otro, la capacidad de comunicación horizontal y membresías de inclusión social. Así, las redes sociales condensan los miedos y las esperanzas colectivas. Se propone analizar las redes sociales desde la sociología cultural en tanto mundos de sentido moral en tensión como alternativa a la perspectiva utilitarista.

Palabras clave: violencia, agresión, acoso, redes sociales, sociología cultural

Sumario: 1. Introducción. 2. Mediaciones tecnológicas. 3. Las redes sociales, desigualdad económica y social. 4. Conclusión. Referencias.

1. Introducción

Las redes sociales -Facebook, Twitter, Instagram, WhatsApp, YouTube, WeChat, entre otras- se han consolidado en

1 Doctor en Sociología. Profesor investigador de la Facultad Latinoamericana de Ciencias Sociales, sede México. Miembro del Sistema Nacional de Investigadores Nivel III. Miembro de la Academia Mexicana de Ciencias. Entre sus últimas publicaciones se encuentran: *Semantics of Violence: Revolt and Political Assassination in Mexico*, Palgrave-MacMillan. *Sociología y Ciencia Ficción: Philip K. Dick más allá de Orwell y Foucault*. Flacso-México (2020). Correo electrónico: nelson.artea@flacso.edu.mx.

los últimos años como medios de comunicación claves para transmitir información, así como debatir social y políticamente en las sociedades contemporáneas (Yar, 2012). A través de ellas las personas expresan sus estados de ánimo con respecto a situaciones diversas, sus opiniones políticas con relación a ciertos acontecimientos, así como su posición sobre cualquier evento social que consideran relevante. Las redes se han convertido en este sentido en una pieza del entramado de producción de la vida social. También son mediaciones tecnológicas a través de las cuales las corporaciones privadas, entidades gubernamentales, partidos políticos y medios de comunicación, difunden mensajes que buscan influir en el comportamiento de las personas (Vishwanath, 2015). No obstante, existen tanto en el espacio de la vida social como política y mediática, actores que a través de las redes difunden actos de violencia y delitos, con lo que pueden alterar o reproducir las brechas de desigualdad y asimetría de poder político y social. Paradójicamente, es gracias a las redes que esas violencias y delitos logran despertar también la indignación y las demandas de reparación civil y penal. Así, el análisis de las redes sociales se caracteriza, por un lado, por facilitar y difundir actos de violencia o la comisión de delitos, pero también permite la construcción de vínculos sociales solidarios, el cuestionamiento del poder o la demanda de acciones en beneficio de personas y grupos tradicionalmente marginados o excluidos.

Existen ciertas manifestaciones de violencia y agresión que se difunden en las redes sociales -amenazas, actos que física o psicológicamente afectan a personas o dañan objetos- que son interpretados como claros ejercicios de poder y dominación que desatan la indignación colectiva. No obstante, las redes sociales han instaurado un nuevo léxico de la violencia que a veces dificulta definir sus contornos, en la medida en que se hace referencias a palabras, acrónimos, imágenes, símbolos e íconos que solo adquieren un carácter amenazante o violento a partir de su presencia en estructuras semánticas específicas en contextos sociales concretos (Abdelzaher, 2019). Al igual que en las relaciones cara a cara, el contexto en el que se intercambian mensajes en las redes sociales establece el ca-

rácter violento de una expresión, una imagen o un acrónimo. Con todo, a diferencia de las relaciones cara a cara, las comunicaciones mediadas por las redes sociales se extienden más allá de su contexto de emisión. De esta forma, un mensaje que no se consideró violento en un momento determinado, lo puede ser poco después o viceversa. De igual forma, un mensaje evaluado como agresivo o violento, puede no perder esta característica en el tiempo o solamente poco a poco.

Algunos trabajos han resaltado que la presencia de expresiones violentas en las redes sociales se debe al anonimato que proporcionan algunas de ellas -como Facebook, WhatsApp o YouTube- garantizando la impunidad del ciber-bullying o el ciber-acoso (Lowry et al, 2016). Sin embargo, otras investigaciones muestran que los ataques anónimos no son más agresivos y violentos que los que se hacen desde cuentas donde explícitamente se conoce a sus emisores (Rost, Stahel and Frey, 2016). Estos últimos pueden ser grupos o individuos que a través de las redes sociales muestran su capacidad potencial o real de depredación frente a sus víctimas. Todo ello con vistas a garantizar su posición de poder o explotación. Pero de nuevo, las redes sociales también se llegan a activar cuando el ciber-bullying o el ciber-acoso generan indignación social, demandando así reparaciones civiles o penales si es el caso.

Caracterizar las redes sociales como herramientas para incrementar o reducir la capacidad de personas, colectivos e instituciones para producir violencia o actos delictivos a través de mensajes, información, infografía, imágenes, íconos o acrónimos, es una forma de aproximarse al tema desde una perspectiva utilitarista de la tecnología. Esta última comprende cualquier innovación técnica a partir de las ventajas comparativas que pueden tener sus usuarios -particularmente ventajas que permiten obtener beneficios específicos a corto y mediano plazo-. Este tipo de interpretación en general considera que la sociedad es un campo de competencia constante por recursos materiales y simbólicos de cualquier tipo, y que, en ese sentido, las mediaciones tecnológicas -como aquellas de las redes sociales- permiten obtener mayores o menores ventajas a los distintos actores sociales en función de las capa-

ciudades y habilidades que tengan para manejar dichas redes, así como los contenidos que coloquen en ellas.

De alguna manera estas interpretaciones derivan en una valoración positiva o negativa de las redes sociales, de la capacidad que tienen para generar nuevas violencias y delitos –en beneficio de las asimetrías de poder y dominación–, así como la oportunidad que representan como medios de empoderamiento o resistencia –alterando de algunas formas las relaciones desiguales que sostienen las brechas de desigualdad política y social–. Así, las redes sociales inspiran las atribuciones sombrías que normalmente se atribuyen a la tecnológica –en tanto racionalidad técnica que despersonaliza–, pero también son consideradas como mediaciones que permiten espacios de autonomía –capaces de impulsar los aspectos más positivos de la organización colectiva–. Las redes sociales conjugan así los referentes culturales del miedo y la esperanza. Por un lado, la racionalización tecnológica de la violencia y el delito; por el otro, la capacidad de comunicación horizontal que genera solidaridad y membresías de inclusión social.

El presente capítulo tiene como objetivo examinar la presencia de estas lógicas en los estudios sobre las redes sociales. En la medida en que se explore de forma crítica cómo se han interpretado las nuevas violencias y delitos desde las redes sociales –en tanto meras herramientas tecnológicas, como hacen las aproximaciones utilitaristas–, será posible sugerir la necesidad de pensarlas también como referentes culturales: como mundos de sentidos morales en permanente tensión. Así, el capítulo propone una primera aproximación para comprender las redes sociales no sólo como un objeto con una cierta materialidad, sino como un objeto cultural que, en tanto tal, está sujeto a la interpretación sobre su sentido en el espacio de la vida pública.

Para desarrollar este argumento, el presente ensayo aborda, en un primer momento, una reflexión en torno a cómo se ha interpretado tradicionalmente la tecnología, particularmente las atribuciones que se le han asignado en tanto externalidad que permite la expansión de los mecanismos de solidaridad social o las relaciones desiguales de poder. A conti-

nuación, se analizan las distintas formas en cómo se atribuyen estas valoraciones positivas y negativas en los estudios de las redes sociales asociadas a la expansión de nuevos delitos y violencias. Finalmente se sugieren algunas primeras reflexiones para pensar las redes sociales en tanto referentes de significación cultural a través de las cuales la sociedad discute la violencia y el delito en tanto fenómenos que evidencian brechas de desigualdad social y política.

2. Mediaciones tecnológicas

En las tradiciones sociológicas que se alimenta del pensamiento de Marx (2000) y Weber (1975), se atribuye al desarrollo tecnológico un cariz particularmente sombrío, capaz de despersonalizar, cosificar, alienar y deshumanizar la vida de las personas. La mecanización -en el caso de una buena parte de los marxismos- fragmenta la vida del hombre a tal punto que se convierte en el engranaje de una gran maquinaria de la producción. La formalización de la vida colectiva por los procesos de clasificación y organización racional de la sociedad -siguiendo a los weberianos-, se cristalizan en la creciente burocracia -en tanto cristalización de una técnica de formalización de la vida- que aplasta la solidaridad colectiva. Ambos autores, pese a todo, no suscribían del todo una visión pesimista de la tecnología. Consideraban que existía la posibilidad de reorientar las fuerzas de esta última por distintas vías: en Marx, a través del control de los medios de producción tecnológica por parte de quienes sufren directamente su dominio; en Weber, mitigando la racionalidad burocrática a través de una vibrante vida política.

Una parte importante de las escuelas de pensamiento que se construyeron a la sombra de estos dos autores -por su puesto que no son los únicos que desarrollan estas reflexiones, pero quizás si los más visibles- estarán constantemente preocupados por resaltar el proceso de racionalización y deshumanizante de la tecnología. Las voces más pesimistas expresan por lo regular su temor a que la tecnología termine por profundizar las brechas de desigualdad y acabe por suprimir

lo que resta de los principios y valores de solidaridad colectiva. Esta fue una discusión particularmente significativa, por ejemplo, cuando se extendió el uso de las computadoras -en la segunda mitad del siglo pasado- y de forma más fuerte cuando las computadoras personales salieron al mercado. El debate sobre el carácter deshumanizante, pero al mismo tiempo liberador del mundo digital se intensificó, sin duda, cuando a principios del siglo XXI, el Internet hizo su aparición en la escena social.

El conjunto de usuarios de la primera generación de Internet, a decir de Castells (1999), estaban inspirados en corrientes utópicas, comunales y libertarias -en algunos casos abiertamente anarquistas- que en ese momento dieron el rostro a lo que posteriormente sería la red de hoy en día. Pese a este impulso de origen, la red terminó por constituir una arquitectura en dos sentidos diferentes. Por un lado, se conformó un grupo de actores que buscaron restringir “el acceso a una minoría de aficionados a la informática, los **únicos** capaces y deseosos de gastar tiempo y energía en vivir en el ciber-espacio” (Castells, 1999, p. 290) y, por otro, se conformó un grupo que expandió el uso del ciber-espacio hacia otros actores, incluidas las grandes corporaciones comerciales y las entidades gubernamentales.²

El principio general que considera que en la actualidad los dispositivos electrónicos conectados al Internet son una expresión amenazante de la tecnología, considera que la computadora convierte cada mensaje, independientemente de su sentido, en una serie de bites y bytes numéricos. Estas últimas están conectadas a otras a través de impulsos eléctricos que se propagan ampliamente a través de redes. Cuando regresan esos bites numéricos al medio de vida humano es interpretado en muchas ocasiones como un proceso que implica la sujeción de la actividad humana a un control racional imper-

2 A decir de Leman-Langlois (2008) el prefijo ciber adquirió una potencia particularmente relevante en el texto de ciencia ficción *Neuromante* de William Gibson. Este autor utilizó dicho prefijo de forma intercambiable para denominar distintas relaciones sociales mediadas tecnológicamente. El término encajó bastante bien en la posterior cultura del Internet en la medida en que permitió subrayar la mediación de las relaciones sociales a través de computadoras y redes de comunicación.

sonal (Alexander, 2003), el cual se caracteriza muchas veces por su violencia y su capacidad de dañar la integridad de las personas.

Esto ha permitido la aparición de nuevas formas de delitos a través de mediación tecnológicas que han sido denominados ciber-delitos, ciber-crímenes (Nham and Huey, 2008), ciber-ataques, ciber-amenazas (Lemieux, 2008), ciber-bullying, ciber-acoso, ciber-violencia (Lowry et al, 2016), ciber-golpeteo digital (Hoffmeister, 2014; Patton, Escherman and Butler, 2013). No obstante, cuando se interpreta que las redes sociales permiten mejorar las condiciones de vida y las relaciones de las personas en la sociedad, los bites numéricos que se transmiten por las redes sociales regresan permitiendo la organización de mecanismos de denuncia, resistencia y movilización contra la violencia, la dominación y los delitos. A estas expresiones adquieren el nombre de ciber-feminismo, ciber-ciudadanías, ciber-pacificadores, los cuales que generan ciber-denuncias o ciber-resistencias colectivas (Nakamura, 2006; Lyon, 2009; Warren, 2015).

Este doble rostro de las redes sociales se aprecia en una serie en los trabajos que muestran la intrincada conexión entre ambos rostros. En la manera en cómo el crimen organizado y pandillas utilizan las redes sociales para mostrar públicamente como coordinan sus acciones o su capacidad para producir violencia. Pero de igual forma es posible observar cómo se articulan las organizaciones vecinales y barriales de forma autónoma y en coordinación con la policía para garantizar la construcción no solo de zonas habitacionales seguras, sino del espacio público en general. Ciertamente este tipo de acciones ha abierto una amplia discusión sobre los efectos que estas acciones en las redes pueden tener: estigmatizar grupos y colectivos que se acusa de provocar actos de violencia o delictivos. Por otro lado, el acoso y la violencia contra las mujeres mediada por las redes sociales permite dar cuenta de procesos complejos de dominación y resistencia. En este sentido, si bien es cierto las redes sociales se han convertido en una herramienta que potencia la capacidad para acosar y violentar a las mujeres, también es cierto que las redes funcionan como

un arma de denuncia y de organización colectiva en contra de victimarios y predadores. Finalmente, hay que sumar al debate la difusión en redes sociales de imágenes y discursos que pueden ser considerados violentos, ofensivos o amenazantes, a las que se llama al mismo tiempo a proscribir y defender en nombre de la libertad de expresión. Esta intrincada forma de moverse de las redes sociales puede examinarse en algunos trabajos empíricos en particular.

2.1 Crimen organizado, pandillas y organización vecinal

El crimen organizado, por ejemplo, utiliza Twitter para comunicar y coordinar sus acciones (Andrews, Brewster and Day, 2018). En algunos casos las bandas organizadas del crimen se muestran en las redes sociales para alardear sobre los hechos delictivos que comenten, lo que les permite mostrar su capacidad futura de acción frente a otros grupos o potenciales víctimas. A este último tipo de actuación se le conoce como “golpeteo digital”. Su intención es mantener viva la presencia de grupos criminales en el imaginario colectivo de una sociedad (Hoffmeister, 2014). De hecho, algunos estudios han mostrado que es posible observar cómo los rumores o la información que se difunde en las redes sociales -entre los miembros de grupos dedicados a actividades delictivas-, pueden ser un adecuado predictor de los ciclos de violencia entre bandas criminales (Moule, Decker and Pyrooz, 2017; Urbanik and Haggerty 2018). Algunos grupos usan Twitter, Facebook y YouTube para intercambiar insultos, amenazas de violencia o para anunciar homicidios o amedrentar a personas y grupos en particular (Patton, Escherman and Butler, 2013). Aunque, ciertamente, no todas las interacciones en las redes sociales entre estos actores terminan irremediabilmente en hechos violentos (Patton et al, 2019).

Si bien grupos del crimen organizado o pandillas se apoyan en las redes, hay también respuesta por parte de la sociedad, a través de distintos mecanismos de autogestión o en coordinación con la policía. Las redes sociales son efectivas para resolver crímenes cuando actúan de forma coordinada comu-

nidades o colectivos con las fuerzas del orden (Beshears, 2017). Al respecto, Facebook y Twitter permiten mejorar ampliamente la predicción del crimen en las áreas urbanas cuando los vecinos de barrios populares y de clase media registran los delitos que se cometen y la ponen a disposición de la policía (Fateh-kia, O'Brien and Weber, 2019). Por otra parte, Twitter permite comprender el ambiente social en barrios, colonias o distritos comerciales en la forma en cómo sus usuarios reportan los hechos delictivos, criminales o la violencia (Malleson and Andresen, 2015). De esta manera es posible medir el ambiente social previo a los actos de violencia a través de las redes sociales, particularmente a través del Twitter (Lan et al, 2019).³

No obstante, también se ha mostrado que la información sobre delitos que se distribuyen vía redes sociales puede generar dinámicas de pánico moral o estigmatización hacia ciertos sectores que tradicionalmente son considerados como potenciales criminales: minorías étnicas, grupos racializados o sectores marginados o excluidos (Kounadi et al, 2015). Por tanto, las redes sociales, advierten varios estudios, pueden, por un lado, construir ciudadanía, comunidad y solidaridad; pero, por otro, refuerzan estereotipos sobre qué es la violencia y sus supuestos perpetradores (Bing, 2015). De esta forma, es conveniente utilizar de manera crítica la información que se puede obtener en las redes, con el fin de comprender la distribución espacial de la violencia, el crimen y el delito, evitando caer en estereotipos que criminalizan a ciertos grupos, incrementando con ello las asimetrías de desigualdad social (Piña-García and Ramírez-Ramírez, 2019).

2.2 Violencia contra las mujeres

Con respecto a la violencia contra las mujeres se puede apreciar también este doble rostro de las redes sociales. Exis-

3 Las redes han mostrado su efectividad para reducir la violencia en otros espacios públicos distintos a los vecindarios, por ejemplo, en los estadios de fútbol. Utilizando las fotografías que los aficionados suben a las redes sociales –a los que suman por lo general distintos comentarios u otras imágenes– se pueden analizar las emociones de estos últimos en tiempo real poco antes de que inicia alguna pelea en entre hinchas (Boychuck et al, 2016).

te un consenso amplio por considerar la ciber-violencia contra las mujeres como aquel abuso cometido de forma repetida contra una mujer usando tecnología digital: llamadas telefónicas amenazantes, estoqueo, monitoreo a través de teléfonos inteligentes o redes sociales (Douglas, Harris and Dragiewicz, 2019).⁴ Así, estas últimas permiten reforzar la violencia contra las mujeres -particularmente, aunque no exclusivamente a través del acoso-, pero también han servido para empoderar a las mujeres, quienes exponen los procesos de estoqueo y monitoreo a los que han sido expuestas, así como los mensajes violentos que reciben, todo lo cual sirve para respaldar las acciones legales contra sus acosadores y predadores (Douglas, Harris and Dragiewicz, 2019).

2.3 Las imágenes de la violencia

El uso de las redes sociales también ha traído una forma distinta de difundir las imágenes de la violencia a tal punto que ha renovado el debate sobre qué tanto influyen dichas imágenes en la producción y reproducción de la violencia entre ciertos sectores de la población -particularmente, niños, adolescentes y jóvenes- y en qué medida dichas imágenes contribuyen a la normalización de la violencia. Hay investigaciones que muestran la fuerte relación que hay entre la exposición a la violencia en las redes sociales y el incremento de las actitudes violentas. Se sugiere que esto se debe a que los usuarios expuestos a esta información e imágenes se insensibilizan ante los hechos de violencia y crueldad (Huesmann, 2007). Particularmente se ha enfatizado que las imágenes violentas que se difunden en redes sociales muestran sobre todo los cuerpos y el sufrimiento de las víctimas, lo que produce que estas últimas sean consideradas como objetos sobre los que se justifica el ejercicio de la violencia (Krupa, 2009). No obstante, hay trabajos que dejan ver que no existe ni teórica, metodológicamente una relación entre la

4 Las redes sociales permiten no solo perpetrar este tipo de violencia digital, sino que el daño es aún mayor cuando se difunden imágenes sexuales o íntimas sin consentimiento, en lo que puede ser tipificado como venganza pornográfica (Henry and Powell, 2016).

violencia que se difunde y consume en las redes y la comisión real de la violencia (Grimes and Bergen, 2008). Este último tipo de trabajos sugieren que hay un conjunto de factores y variables intermedias que es necesario explorar para poder comprender hasta qué punto la difusión de la violencia en las redes influye en la creación de violencia.

Desde estas interpretaciones la violencia en las redes -incluso la que es posible encontrar en los juegos en línea-, así como una parte importante de imágenes de contenido sexual explícito en ellas deben ser consideradas como expresiones que definen “zonas de excepción cultural”, donde la gente puede explorar fantasías antisociales y violentas sin afectar la vida o la integridad de las personas (Atkinson and Rodgers, 2015). Para este tipo de trabajos resulta fundamental establecer hasta qué punto las expresiones de rechazo y censura de las imágenes, discursos y narrativas en las redes sociales, expresan la agenda de los medios o grupos conservadores para generar pánico moral, en nombre de “proteger o salvaguardar” la conciencia colectiva, las “buenas costumbres” y los “valores tradicionales” (Cricher, 2012).

En contraposición de esta literatura existe otra que advierte que más allá de la defensa de las supuestas buenas conciencias, las redes sociales terminan por reproducir y legitimar los discursos violentos contra grupos considerados vulnerables -sobre todo, grupos indígenas, minorías raciales, infantes, adolescentes o mujeres-, garantizando la reproducción de la violencia simbólica al naturalizar los sistemas de dominación (Recuero, 2015). En este sentido, quienes a veces demandan la defensa de las expresiones de violencia de distinto tipo en las redes sociales -resguardándose en el derecho a la libertad de expresión- ignoran que ello es una oportunidad también para los discursos de odio de extremistas conservadores o de izquierda (Cambron, 2019).

3. Las redes sociales, desigualdad económica y social

Si bien es cierto, muchas de las narrativas visuales y discursivas en las redes sociales terminan por reproducir las

estructuras de dominación y poder, en otros casos dichas narrativas detonan la movilización social y las demandas de reparación civil (Roberts and Marchais, 2018). Así, la exposición de actos de violencia y amenaza a través de las redes sociales desatan lo que se ha denominado como “tormentas de fuego cibernéticas” contra cualquiera de sus perpetradores, sean estas empresas, políticos, celebridades, medios de comunicación o académicos; por tanto, resultan relevantes porque convierten hechos indignantes en profundos y amplios debates públicos (Rost, Stahel and Frey, 2016). De esta forma, las redes sociales han cambiado la forma en cómo se captura la violencia, cómo se reporta, se analiza y se actúa en consecuencia contra ella, y no sólo en función de su reproducción según las relaciones desiguales de poder en una sociedad determinada.

Por ejemplo, hay estudios que muestran que aquellos eventos de violencia que se difunden por las redes sociales, en los que se confrontan minorías sociales o políticas contra fuerzas del orden, está directamente relacionado con el grado en el que se comprometen los activistas sociales. Así, entre más se difundan las imágenes –tanto extensiva como intensivamente– de abusos de policías y militares en las redes sociales contra ciudadanos o movimientos sociales, se aprecia un mayor compromiso político y un mayor compromiso social en defensa de las víctimas, los valores democráticos y los principios de inclusión ciudadana (Masías, Hecking and Hoppe, 2018; Bejan et al, 2018).

La reflexión sobre la doble cara de las redes sociales, tanto los nuevos tipos de violencia y delito que producen, así como las nuevas formas de demandas de reparación civil y solidaridad que le acompañan, no debe entenderse solamente como las posibilidades instrumentales que representan. Las redes sociales son algo más que una herramienta que se puede utilizar para fines positivos o negativos, a veces moralmente aceptables, otras censurables. Las redes sociales son también mundos de sentido, conexiones de interacción y relaciones simbólicas que se operan y ponen en juego a través del intercambio de ideas, imágenes, textos, discursos e íconos. Son, por tanto, una red cultural que, como entramado de códigos, con-

densan las aspiraciones, los sentimientos y las emociones de actores diversos; en última instancia, se instituyen en discursos que permiten dar sentido al mundo social. Las narrativas que emergen de las redes sociales no sólo refieren a marcos de acción específicos, sino que son formas de construir *pathos* -un sentimiento o experiencia- sobre la organización criminal, las pandillas, la violencia, las amenazas, el acoso contra las mujeres, la organización vecinal, los colectivos en resistencia, la creación artística, la cooperación o confrontación con el poder público, las corporaciones o partidos políticos.

La difusión de estos *pathos* en las redes sociales genera discusiones constantes en la vida pública, en donde se cuestiona la legitimidad de la violencia y la comisión de delitos, y se establecen regulaciones para sanción y la protección de las víctimas. En este sentido, las redes sociales están constantemente sujetas a interpretación y crítica sobre lo que expresan, proyectan o representan. Es en estas disputas, que se llama a veces a reducir las brechas de desigualdad social. Los actos de violencia hacia ciertos sectores de la población -mujeres, jóvenes, niños, indígenas, grupos racializados, marginados o excluidos- por parte otros grupos con capacidad para producir violencia tanto de forma individual, organizada o institucionalizada son expuestos en las redes para mostrar el poder desmedido con el que cuentan. Estos actos de violencia generan al mismo tiempo respuestas tanto de apoyo como de solidaridad. Es en estas disputas que las brechas de desigualdad económica y de poder quedan en evidencia y desatan confrontaciones en la esfera civil y política sobre cuales deben ser las bases que dan sustento a la vida de una sociedad, así como las exigencias de cambio social. Incluso en algunas ocasiones dichas confrontaciones ponen en evidencia brechas de desigualdad económica y de poder que hasta ese momento no se habían percibido como tales.

Las redes sociales permiten construir espacio de discusión pública sobre las desigualdades sociales de una forma en la que antes no era posible. En dichas discusiones se ponen en juego una serie de valores y principios normativos que al mismo tiempo apuestan por dilatar o expandir los mecanismos de solidaridad o las membresías de inclusión ciudadana. Cierta-

mente no todos los temas que cruzan las redes sociales provocan algún tipo de debate. Esto último depende en muchas de las ocasiones del tipo de preocupación, valores o principios que se movilizan en ellas. Se podría argumentar al respecto que existe también una brecha de desigualdad en el acceso a estas mediaciones electrónicas. Esto es cierto para el caso de América latina, donde hay una distribución desigual de las tecnologías de la información. Si bien en Chile o Argentina el 66.5% y el 59.9% de las personas tienen acceso a los servicios de Internet móvil respectivamente, en Honduras o El Salvador menos del 25% lo tienen. No obstante, el porcentaje de hogares conectados a una red de Internet es mucho menor en el conjunto de la región. No son muchos los países que rebasan el 50% (Banco Mundial, 2015). Esto marca una tendencia particularmente preocupante en la región, donde los debates que se producen en la red no necesariamente son accesibles para todos.

No obstante, aun cuando las brechas digitales se reduzcan -que al parecer así está sucediendo-, lo cierto es que eso tampoco redundará en una mayor participación en los debates que se generen en las redes sociales en torno a como la violencia y el delito exponen y expresan las condiciones de marginación, desigualdad y exclusión social. Las redes son, como se ha dicho aquí, un espacio en el que se narrativizan distintos *pathos* sobre hechos de violencia y crimen. El camino para construir ese *pathos* no está sujeto a ningún determinismo tecnológico, es un camino en el que discursos, imágenes o íconos se articulan para tratar de producir una reacción colectiva.

Para ello se requieren actores colectivos que signifiquen las condiciones de explotación y dominación, así como las asimetrías de poder que afecta a ciertos grupos sociales (Conner-ton, 1989). Dichos actores pueden ser considerados como grupos portadores (Eisenstadt, 1982; Eyerman and Jamison, 1990), agentes que operan como portavoces de víctimas o personas afectadas por alguna injusticia o daño. Estos pueden ser elites o clases marginales, líderes religiosos, parias espirituales, líderes de opinión, intelectuales o artistas. Los grupos portadores traducen el reclamo de las víctimas en discursos y narrativas persuasivas a través de recursos simbólicos, y los llevan a una

esfera pública a través de las redes sociales a una sociedad ciertamente fragmentada y diferenciada. En este sentido, si se quiere generar un debate sobre las desigualdades y las brechas de poder y desigualdad en las sociedades contemporáneas es necesario reflexionar en la necesidad de producir grupos portadores que sean capaces de transcribir las demandas colectivas que permitan expandir la solidaridad y la inclusión social.

4. Conclusión

Las redes sociales son mediaciones tecnológicas a través de las cuales se disputa lo que se considera una amenaza y oportunidad para la solidaridad social; aquello que se considera una intrusión potencialmente destructiva o reparadora de las membresías de inclusión social. Cuando se expone la violencia a través de las redes sociales, el acoso, el crimen y otras formas de explotación y dominación social, también emergen a través de ellas las demandas de cohesión, de sanción y de restablecimiento de los vínculos sociales horizontales. Se exige que los victimarios sean castigados y las víctimas reconocidas por las estructuras jerárquicas de orden institucional. Incluso, se pide que se apliquen las sanciones correspondientes si se han afectado el interés y la dignidad de las personas, sobre todo si las víctimas se encuentran en una posición de desigualdad, desventaja y minusvalía social. Dichas demandas, sin embargo, no emergen de manera automática, son el resultado de interpretaciones críticas que se hacen desde distintos espacios sociales y no necesariamente están exentos de resistencias. Siempre hay actores que no reconocen los actos de violencia, acoso, crimen o abuso como tales, ni mucho menos consideran que existen procesos de desigualdad social que los exacerban. Ese diferencial interpretativo abre disputas por el control del sentido de lo que aparece en las redes sociales. No obstante, dicha disputas evidencian, independientemente de su resultado, una ventana a través de la cual se dejan ver las tensiones sobre el sentido de la violencia en las redes sociales.

Considero que analizar las redes sociales y su relación con la violencia como un espacio de competencia por controlar el sentido de lo que en ellas circula es un primer paso para

examinar dichas redes más allá de la tradicional aproximación utilitaria que las aborda como una mera externalidad técnica, sujeta a su buen o mal manejo. Las redes sociales y la violencia pueden ser apreciadas como parte de un sistema-signo donde los actores tratan de clasificar qué es un riesgo y qué es una oportunidad para generar solidaridad social. Si la propuesta que se ha hecho aquí, de forma preliminar, desde la sociología cultural puede resultar sugerente, es porque las redes sociales y la violencia condensan las disputas sociales sobre la existencia del mal y su relación conflictual con el bien.

Como sugirió Geertz (2003) hace tiempo, el mal es todo aquello que una sociedad coloca en una cosmovisión determinada como “la verdadera naturaleza de las fuerzas destructivas que moran en la persona y fuera de él, de tal manera que sea posible llegar a una especie de acuerdo con esos fenómenos” (Geertz, 2003, p. 121). El mal regularmente no es tratado en las sociedades como una fuerza irreal, más bien se le acepta y se caracteriza “y cierta actitud frente al mal -resignación, oposición activa, evasión hedonista, recriminación a uno mismo y arrepentimiento o humildes súplicas para obtener misericordia- es considerada razonable y apropiada dada su naturaleza” (Geertz, 2003, p. 121). De esta forma lo que una sociedad tematiza y considera como una expresión de desorden está de alguna manera simbolizado y dramatizado en distintos ámbitos de la vida social y no se encuentran por supuesto fuera de su sistema cultural. De alguna manera las redes sociales y la violencia apelan constantemente a ese desorden de forma dramatizada una y otra vez. Por ello se hace necesario comprenderlas en su interrelación con el fin entender cuáles son las emociones y los sentimientos, las esperanzas y los medios, que se anidan en la sociedad contemporáneas.

Referencias

- Abdelzaher, E. M. (2019) ‘Lexicon-based Detection of Violence on Social Media’, *Cognitive Semantics*, 5(1), pp. 32-69 [online]. Available at: <https://doi.org/10.1163/23526416-00501002>.

- Alexander, J. C. (2003) *The meanings of social life: A cultural sociology*. New York: Oxford University Press.
- Andrews, S., Brewster, B. and Day, T. (2018) 'Organised crime and social media: A system for detecting, corroborating and visualising weak signals of organised crime online', *Security Informatics*, 7(1), p. 3 [online]. Available at: <https://doi.org/10.1186/s13388-018-0032-8>.
- Atkinson, R. and Rodgers, T. (2015) 'Pleasure Zones and Murder Boxes: Online Pornography and Violent Video Games as Cultural Zones of Exception', *British Journal of Criminology*, 56(6), pp. 1291-1307 [online]. Available at: <https://doi.org/10.1093/bjc/azv113>.
- Banco Mundial (2015) 'Indicadores del desarrollo mundial'. *Banco Mundial* [online]. Available at: http://www.google.com/publicdata/explore?ds=d5bncppjof8f9_#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2&scale_y=lin&ind_y=false&rdim=region&idim=country:HND:SLV:ARG:BOL:BRA:CHL:COL:MEX:PER:URY&idim=region:ECS:SSF:LCN:NAC&ifdim=region&tstart=1122958800000&tend=1375419600000&hl=es&dl=es&ind=false.
- Bejan, V., Hickman, M., Parkin, W. S. and Pozo, V. F. (2018) 'Primed for death: Law enforcement-citizen homicides, social media and retaliatory violence'. *PLOS ONE*, 13(1), e0190571 [online]. Available at: <https://doi.org/10.1371/journal.pone.0190571>.
- Beshears, M. L. (2017) 'Effectiveness of Police Social Media Use'. *American Journal of Criminal Justice*, 42(3), pp. 489-501 [online]. Available at: <https://doi.org/10.1007/s12103-016-9380-4>.
- Bing, N. (2015) 'Kenya Decides: Kiswahili, social media and politics in Kenya's 2013 general elections', *Journal of African Media Studies*, 7(2), pp. 165-183 [online]. Available at: https://doi.org/10.1386/jams.7.2.165_1.

- Boychuk, V., Sukharev, K., Voloshin, D. and Karbovskii, V. (2016) 'An Exploratory Sentiment and Facial Expressions Analysis of Data from Photo-sharing on Social Media: The Case of Football Violence', *Procedia Computer Science*, 80, pp. 398-406 [online]. Available at: <https://doi.org/10.1016/j.procs.2016.05.340>.
- Cambron, R. J. (2019) 'World War Web: Rethinking "Aiding and Abetting" in the Social Media Age', *Case Western Reserve Journal of International Law*, 51(1), pp. 293-325.
- Castells, M. (1999) *La era de la información. Economía, sociedad y cultura: La sociedad red*. México, DF: Siglo XXI.
- Connerton, P. (1989) *How Societies Remember*. Cambridge: Cambridge University Press.
- Critcher, C. (2012) 'Screen Savers. Case Histories of Social Reaction to Mass Media, Children and Violence', *Recherches Sociologiques et Anthropologiques*, 43(1), p. 59-78 [online]. Available at: <https://doi.org/10.4000/ras.841>.
- Douglas, H., Harris, B. A. and Dragiewicz, M. (2019) 'Technology-facilitated Domestic and Family Violence: Women's Experiences', *The British Journal of Criminology*, 59(3), pp. 551-570 [online]. Available at: <https://doi.org/10.1093/bjc/azy068>.
- Eisenstadt, S. (1982) 'The Axial Age: The Emergence of Transcendental Visions and the Rise of Clerics', *European Journal of Sociology*, 23(2), pp. 294-314 [online]. Available at: <https://doi.org/10.1017/S0003975600003908>.
- Eyerman, R. and Jamison A. (1990) *Social Movements: A cognitive Approach*. Londres: Polity.
- Fatehkia, M., O'Brien, D. and Weber, I. (2019) 'Correlated impulses: Using Facebook interests to improve predictions of crime rates in urban areas', *PLOS ONE*, 14(2), e0211350 [online]. Available at: <https://doi.org/10.1371/journal.pone.0211350>.
- Geertz, C. (2003) *La interpretación de las culturas*. Barcelona: Gedisa.

- Grimes, T. and Bergen, L. (2008) 'The Epistemological Argument Against a Causal Relationship Between Media Violence and Sociopathic Behavior Among Psychologically Well Viewers', *American Behavioral Scientist*, 51(8), pp. 1137-1154. [online]. Available at: <https://doi.org/10.1177/0002764207312008>.
- Henry, N. and Powell, A. (2016) 'Sexual Violence in the Digital Age: The Scope and Limits of Criminal Law', *Social & Legal Studies*, 25(4), pp. 397-418 [online]. Available at: <https://doi.org/10.1177/0964663915624273>.
- Hoffmeister, T. (2014) 'The Challenges of Preventing and Prosecuting Social Media Crimes', *Pace L. Rev*, 35(1), pp. 115-134.
- Huesmann, L. R. (2007) 'The Impact of Electronic Media Violence: Scientific Theory and Research'. *Journal of Adolescent Health*, 41(6), pp. S6-S13 [online]. Available at: <https://doi.org/10.1016/j.jadohealth.2007.09.005>.
- Kounadi, O., Lampoltshammer, T. J., Groff, E., Sitko, I. and Leitner, M. (2015) 'Exploring Twitter to Analyze the Public's Reaction Patterns to Recently Reported Homicides in London', *PLOS ONE*, 10(3), e0121848 [online]. Available at: <https://doi.org/10.1371/journal.pone.0121848>.
- Krupa, C. (2009) 'Histories in red: Ways of seeing lynching in Ecuador', *American Ethnologist*, 36(1), pp. 20-39 [online]. Available at: <https://doi.org/10.1111/j.1548-1425.2008.01107.x>.
- Lan, M., Liu, L., Hernandez, A., Liu, W., Zhou, H. and Wang, Z. (2019) 'The Spillover Effect of Geotagged Tweets as a Measure of Ambient Population for Theft Crime', *Sustainability*, 11(23), 6748 [online]. Available at: <https://doi.org/10.3390/su11236748>.
- Leman-Langlois, S. (2008) 'Introduction: Technocrime' in Leman-Langlois, S. (ed.) *Technocrime: Technologie, crime and social control*. Londres: Willan Publishing, pp. 1-12.
- Lemieux, F. (2008) 'Information technology and criminal intelligence a comparative perspective' in Leman-Langlois, S. (ed.) *Technocrime: Technologie, crime and social control*. Londres: Willan Publishing, pp. 139-168.

- Lowry, P. B., Zhang, J., Wang, C. and Siponen, M. (2016) 'Why Do Adults Engage in Cyberbullying on Social Media? An Integration of Online Disinhibition and Deindividuation Effects with the Social Structure and Social Learning Model', *Information Systems Research*, 27(4), pp. 962-986 [online]. Available at: <https://doi.org/10.1287/isre.2016.0671>.
- Lyon, D. (2009) *Identifying citizens: ID cards as surveillance*. Londres: Polity.
- Malleson, N. and Andresen, M. A. (2015) 'The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns', *Cartography and Geographic Information Science*, 42(2), pp. 112-121. [online]. Available at: <https://doi.org/10.1080/15230406.2014.905756>.
- Marx, K. (2000) *El manifiesto del partido comunista*. Buenos Aires: Ediciones El Aleph.
- Masías, V. H., Hecking, T. and Hoppe, U. (2018) 'Social networking site usage and participation in protest activities in 17 Latin American countries', *Telematics and Informatics*, 35(7), pp. 1809-1831 [online]. Available at: <https://doi.org/10.1016/j.tele.2018.05.010>.
- Moule, R. K., Decker, S. H. and Pyrooz, D. C. (2017) 'Technology and conflict: Group processes and collective violence in the Internet era'. *Crime, Law and Social Change*, 68(1-2), pp. 47-73 [online]. Available at: <https://doi.org/10.1007/s10611-016-9661-3>.
- Nakamura. L. (2006) 'Cultural difference, theory, and cyber-culture Studies: A case of mutual repulsion' in Siver, D. and Massanari, A. (eds.) *Critical Cyber Culture Studies* New York: New York University Press, pp. 29-36.
- Nham, J. and Huey, L. (2008) 'Policing through nodes, clusters and bandwidth' in Leman-Langlois, S. (ed.) *Technocrime: Technologie, crime and social control*. Londres: Willan Publishing, pp. 66-87.

- Patton, D. U., Eschmann, R. D. and Butler, D. A. (2013) 'Internet banging: New trends in social media, gang violence, masculinity and hip hop', *Computers in Human Behavior*, 29(5), pp. A54-A59 [online]. Available at: <https://doi.org/10.1016/j.chb.2012.12.035>.
- Patton, D. U., Leonard, P., Elaesser, C., Eschmann, R. D., Patel, S. and Crosby, S. (2019) 'What's a Threat on Social Media? How Black and Latino Chicago Young Men Define and Navigate Threats Online', *Youth & Society*, 51(6), pp. 756-772 [online]. Available at: <https://doi.org/10.1177/0044118X17720325>.
- Piña-García, C. A. and Ramírez-Ramírez, L. (2019) 'Exploring crime patterns in Mexico City', *Journal of Big Data*, 6(1), p. 65 [online]. Available at: <https://doi.org/10.1186/s40537-019-0228-x>.
- Recuero, R. (2015) 'Social Media and Symbolic Violence'. *Social Media + Society*, 1(1), 205630511558033 [online]. Available at: <https://doi.org/10.1177/2056305115580332>.
- Roberts, T. and Marchais, G. (2018) 'Assessing the Role of Social Media and Digital Technology in Violence Reporting', *Contemporary Readings in Law and Social Justice*, 10(2), pp. 9-42 [online]. Available at: <https://doi.org/10.22381/CRLSJ10220181>.
- Rost, K., Stahel, L. and Frey, B. S. (2016) *Digital Social Norm Enforcement: Online Firestorms in Social Media* [Data set]. Available at: <https://doi.org/10.1371/journal.pone.0155923>.
- Urbanik, M.-M. and Haggerty, K. D. (2018) "'#It's Dangerous': The Online World of Drug Dealers, Rappers and the Street Code', *The British Journal of Criminology*, 58(6), pp. 1343-1360 [online]. Available at: <https://doi.org/10.1093/bjc/azx083>.
- Vishwanath, A. (2015) 'Diffusion of deception in social media: Social contagion effects and its antecedents', *Information Systems Frontiers*, 17(6), pp. 1353-1367 [online]. Available at: <https://doi.org/10.1007/s10796-014-9509-2>.

- Warren, T. C. (2015) 'Explosive connections? Mass media, social media, and the geography of collective violence in African states', *Journal of Peace Research*, 52(3), pp. 297-311 [online]. Available at: <https://doi.org/10.1177/0022343314558102>.
- Weber, M. (1975) *Economía y sociedad*. México DF: Fondo de Cultura Económica.
- Yar, M. (2012) 'E-Crime 2.0: The criminological landscape of new social media', *Information & Communications Technology Law*, 21(3), pp. 207-219 [online]. Available at: <https://doi.org/10.1080/13600834.2012.744224>.

INTELIGÊNCIA ARTIFICIAL GENERATIVA FRENTE AO DEVER DE FUNDAMENTAR AS DECISÕES JUDICIAIS: O CHATGPT E OS LIMITES DO ARTIGO 489, §1º, DO CPC

Darci Guimarães Ribeiro¹
Afonso Vinício Kirschner Fröhlich²
DOI: 10.29327/5312711.1-7

Resumo: Muito embora o tema do emprego da Inteligência Artificial (IA) no processo judicial já esteja há anos sendo discutido pelos juristas, o final de 2022 e início de 2023 trouxeram ao debate público a IA Generativa. Trata-se da possibilidade de modelos que produzem textos coesos e completos, como se construídos por humanos, sendo seu maior exemplo a sensação *ChatGPT*. O impacto dessa nova ferramenta é tamanho que já se cogita a utilização em decisões judiciais, o que faz surgir uma necessidade de ressignificar o dever de fundamentação, consabidamente caro ao Estado Democrático de Direito. Nesse cenário incerto, o presente artigo enfrenta o seguinte problema de pesquisa: em que medida a Inteligência Artificial

1 Advogado e Árbitro. Pós-Doutor em Direito Processual Constitucional pela *Università degli Studi di Firenze*. Doutor em Direito pela *Universitat de Barcelona*. Especialista e Mestre pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Professor Titular de Direito Processual Civil da Unisinos e PUCRS. Professor Adjunto de Direito Processual Civil da UFN. Professor do Programa de Pós-Graduação em Direito da Unisinos (Mestrado, Doutorado e Pós-Doutorado). Membro da *International Association of Procedural Law*. Membro do Instituto Ibero-americano de Derecho Procesal. Membro do Instituto Brasileiro de Direito Processual Civil. Membro da Academia Brasileira de Direito Processual. Membro da Asociación Argentina de Derecho Procesal. Huésped de Honor da Casa de Altos Estudios da Universidad Nacional de Córdoba, Argentina. Membro representante do Brasil no Projeto Internacional de Investigação sobre *Nuevos retos tecnológicos del derecho probatorio*, (PID2020-115304GB-C21) financiado pelo Ministério de Ciência e Inovação – MEC – da Espanha. E-mail: darci.guimaraes@terra.com.br.

2 Mestre em Direito Público pela Universidade do Vale do Rio dos Sinos (Unisinos). Bacharel em Direito pela mesma Universidade. Especialista em Direito e Negócios Imobiliários pela Fundação Escola Superior do Ministério Público (FMP). Advogado sócio do Escritório de Advocacia Afonso Fröhlich Advogados Associados. E-mail: afonsovinicio@afrohlich.adv.br e afonsovkf@gmail.com.

Generativa impacta na fundamentação das decisões judiciais, levando em consideração os limites estabelecidos pelo artigo 489, §1º, do CPC? Com base nesse questionamento e utilizando-se da metodologia fenomenológica-hermenêutica, o artigo objetiva: (a) contextualizar o atual estágio de desenvolvimento da Inteligência Artificial e apresentar a ferramenta *ChatGPT*; (b) analisar a experiência brasileira na utilização de Inteligência Artificial no âmbito do processo judicial; (c) estudar algumas hipóteses de utilização da IA Generativa no mundo jurídico; e (d) verificar em que medida ferramentas de IA Generativa, especialmente o *ChatGPT*, impactam na fundamentação das decisões judiciais. Ao final, conclui-se que a utilização da IA Generativa como ferramenta auxiliar na decisão deverá estar em compasso com os limites previstos no art. 489, §1º, do CPC, sem os quais não há fundamentação legítima e válida.

Palavras-chave: Inteligência Artificial. *ChatGPT*. Processo. Decisão Judicial. Fundamentação.

Sumário: 1. Introdução. 2. O estado da arte da Inteligência Artificial e as novas ferramentas disponíveis no mercado. 3. A experiência brasileira com a utilização da Inteligência Artificial no Processo Judicial. 4. A utilização da IA Generativa no mundo jurídico. 5. ChatGPT e a fundamentação das decisões judiciais. 6. Considerações finais (e uma experiência). Referências

1. Introdução

A trajetória de estudos que culminou na utilização da Inteligência Artificial (IA) no Direito é de longa data³, sendo hoje perceptível sua aplicação em tarefas jurídica cotidianas, nos escritórios e serventias. Apesar da caminhada de anos, desde o final de 2022 (e, principalmente, nos primeiros meses de 2023), a discussão sobre a extensão da abertura sistêmica

3 Sobre alguns importantes eventos precursores à era da revolução digital, remeta-se ao recente artigo de Antônio Pereira Gaio Júnior e Fábiana Antonio Silva (2023), intitulado *Direito, Processo e Inteligência Artificial: diálogos necessários ao exercício da jurisdição*.

à essa nova tecnologia passou a ser acentuada ainda mais pela disseminação de uma poderosa e revolucionária nova ferramenta. Por isso não seria exagero dizer que estamos entrando na Pós-humanidade, que é a ideologia da IA, nas palavras de Pérez-Luño (2020)⁴.

Trata-se do *ChatGPT*, um sistema de Inteligência Artificial Generativa, que não foi criado especificamente para atuar no mundo jurídico, diferente de outras ferramentas já bastante conhecidas⁵, porém que poderia ter sido, tamanho é o impacto que pode vir a ter. Ao menos, possibilidades neste sentido têm sido estudadas e defendidas, com argumentos consistentes para ambos os lados.

Fato é que essa nova ferramenta reacende velhos e incentiva novos debates e questionamentos sobre as possibilidades de utilização da Inteligência Artificial no processo judicial. Ao contrário de ser divagação, essa possibilidade não parece destemperada, principalmente porque a ferramenta é capaz de criar textos coesos e completos, assemelhando-se à redação humana, ou mesmo dela sendo indistinguível. Afinal, não há mais dúvidas de que o *ChatGPT* já alcança êxito no prestigiado “Teste de Turing”, decidindo (aparentemente) como se humano fosse⁶.

Com olhares voltados a esta realidade que vem sendo apresentada e descoberta pelo jurista a cada dia, este artigo enfrenta o seguinte problema de pesquisa: em que medida a Inteligência Artificial Generativa impacta na fundamentação das decisões judiciais, levando em consideração os limites estabelecidos pelo artigo 489, §1º, do CPC?

Para isso, ao enfrentar esse complexo questionamento, o presente artigo objetiva: (a) contextualizar o atual estágio de desenvolvimento da Inteligência Artificial e apresentar a fer-

4 De acordo com a correta opinião de Pérez-Luño, “*el posthumanismo no implica la mejora, perfeccionamiento o actualización del legado humanista, sino que supone su negación, abolición o suplantación*” (Pérez-Luño, 2020, p. 18).

5 Em 2016, por exemplo, ocorreu o anúncio pela IBM do denominado “primeiro robô advogado do mundo”, apelidado de Ross, que fora “contratado” (expressão utilizada extensivamente pela mídia internacional) pelo Escritório de Advocacia americano Baker & Hostetler (Jesus, 2016).

6 Acerca deste tema, consultar a obra *The most Human Human* de Brian Christian (2011).

ramenta *ChatGPT*; (b) analisar a experiência brasileira na utilização de Inteligência Artificial no âmbito do processo judicial; (c) estudar algumas hipóteses de utilização da IA Generativa no mundo jurídico; e (d) verificar em que medida ferramentas de IA Generativa, especialmente o *ChatGPT*, impactam na fundamentação das decisões judiciais, a partir dos limites estabelecidos na legislação processual civil.

A metodologia utilizada é a fenomenológico-hermenêutica, pela qual o pesquisador busca ultrapassar a clássica cisão entre sujeito e objeto, estando inserido e sofrendo os efeitos da realidade do seu objeto de estudo (a Inteligência Artificial em geral e o *ChatGPT* em específico, ambos aplicados ao processo judicial), atribuindo a ela sentido jurídico. Com base nesta metodologia, poder-se-á propor algumas respostas de verniz teórico e prático, tão necessárias frente à novidade que se apresenta e as discussões que dela advêm.

Portanto, é este artigo dividido nas seguintes seções: (1) O estado da arte da Inteligência Artificial e as novas ferramentas disponíveis no mercado; (2) A experiência brasileira com a utilização da Inteligência Artificial no processo judicial; (3) A utilização da IA Generativa no mundo jurídico; e (4) *ChatGPT* e a fundamentação das decisões judiciais. Ao concluir, serão apresentadas algumas considerações finais, bem como uma singela experiência realizada pelos autores deste artigo por curiosidade acadêmica: a resposta do próprio *ChatGPT* à pergunta-problema que norteou este trabalho.

2. O estado da arte da Inteligência Artificial e as novas ferramentas disponíveis no mercado

Como já é de amplo conhecimento, pelo seu impacto em diversas áreas e profissões, os avanços em matéria de Inteligência Artificial permitiram “*crear máquinas que interactúen com nosotros en el terreno de las cualidades humanas*” (Latorre, 2019, p. 87). Essa definição de máquinas inteligentes proposta pelo professor espanhol José Ignacio Latorre (2019), embora não seja técnica e cientificamente completa, é relevante para evidenciar qual é, ao fim e ao cabo, o objetivo do desenvolvi-

mento da IA: atuação em tarefas específicas, com qualidades humanas. Ou, como prefere Marvin Minsky (1968, p. 23): “*artificial intelligence is the science of making machines do things that would require intelligence if done by men*”.

A IA muito se desenvolveu nesse sentido nos último cinco a dez anos. Tanto que é hoje utilizada em âmbitos tão profundos e cotidianos que passa despercebida, com seus usuários sequer sabendo que estão diante de sistemas inteligentes. É o caso, para ilustrar, (i) das redes sociais, que utilizam IA não só para direcionamentos e indicações como para moderação de conteúdos abusivos e para formação de perfis de consumo; (ii) dos transportes; (iii) dos serviços de entrega (*delivery*); (iv) da previsão do tempo; (v) do reconhecimento de imagens; dentre vários outros exemplos citados por Vladmir Aras (2020).

Principalmente a partir do meio para o final de 2022, teve início uma nova fase do estado da arte da Inteligência Artificial. Foi quando passaram a estar disponíveis, para acesso amplo e irrestrito, novos modelos de *Inteligência Artificial Generativa (Generativa AI)*, que são aqueles nos quais redes neurais artificiais podem criar conteúdos e dados novos, partindo dos dados utilizados em seu treinamento, já disponíveis em seu *dataset* (Mantegna, 2020). E o grande exemplo, que circulou e vem circulando os canais de comunicação mundiais nos últimos meses, apesar de não ser o único, é o ChatGPT que, inclusive, permite a contratação da versão *Plus*, já com acesso ao GPT4.

Desenvolvido pela empresa *Open IA*, voltada à pesquisa e criação de modelos de Inteligência Artificial para benefício humano (*About*, 2023?), o modelo denominado ChatGPT interage com seres humanos, em uma conversação que utiliza linguagem natural. De acordo com a empresa desenvolvedora “*The dialogue format makes it possible for ChatGPT to answer follow-up questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests*” (*Introducing ChatGPT*, 2023?).

Em suma, trata-se de um modelo de linguagem em grande escala, treinado com uma quantidade enorme de dados da *internet*, capaz de responder prontamente a qualquer questionamento, podendo escrever textos, poemas, letras de músi-

cas, códigos de computador, estratégias de vendas e, inclusive, resolver problemas complexos, como de física e matemática (Lubowitz, 2023). Para Salvagno, Taccone e Gerli (2023), Chatbots como o ChatGPT utilizam algoritmos programados para compreender os *inputs* (dados inseridos em linguagem natural) e responder a questionamento de forma apropriada, seja com textos pré-redigidos seja com textos criados pela própria IA. O funcionamento e a utilização prática da ferramenta são bem explicados pelos pesquisadores:

ChatGPT is constantly improved with reinforcement techniques, natural language processing, and machine learning, to improve its ability to understand and thoroughly respond to users' needs. Concretely, you can conversationally ask anything and receive a fast and adequate human-like written reply to your questions or requests such as: (a) write a small text on a given topic; (b) get information on a topic of interest; (c) compose an email or message with a certain tone, specific content, and intended for a particular person; (d) correct the shape of a text or change its wording; (e) solve problems. (Salvagno, Taccone and Gerli, 2023).

Desmistifica-se, ao menos parcialmente, a relevância que a ferramenta vem recebendo. Trata-se de sistema com flagrante aplicação prática, com fácil acesso e diálogo ainda mais simples, permitindo uma profunda interação por meio de perguntas e respostas, sobre inumeráveis assuntos, dos cotidianos aos mais complexos. Sobre isso, aliás, parece ser evidente a comparação com o “Teste de Turing”⁷, que também se desenvolve a partir de uma dinâmica de perguntas e respostas. Para que o modelo de IA tenha êxito no teste do matemático e cientista de dados britânico, um interrogador humano, após propor algumas perguntas à máquina, não conseguir descobrir se as respostas são dadas por pessoa ou de um computador (Russel and Norvig, 2013, p. 25).

Apesar das suas fascinantes potencialidades (e certamente por elas), modelos de IA generativa vêm levantando uma série de discussões envolvendo principalmente direito de autoria, plágio, inconsistências e erros nas respostas. H.

7 Trata-se de uma dinâmica elaborada por Alan Turing na metade do século XX para responder ao questionamento se poderiam as máquinas pensar, a partir de suas capacidades para mimetizar o cérebro humano (Turing, 1950).

Holden Thorp, editor-chefe da revista *Science*, por exemplo, em janeiro de 2023, publicou editorial descrevendo algumas das sérias implicações na ciência e na academia de programas de geração de textos com o *ChatGPT*. Para ele, um programa de Inteligência Artificial não pode ser autor e, em razão disso, diz estar atualizando as políticas editoriais das publicações do Grupo *Science* para especificar que textos gerados por *ChatGPT* não são autorais e não serão aceitos (Thorp, 2023).

Nesse mesmo sentido, no início de 2023, a *International Conference on Machine Learning* publicou uma nota proibindo artigos com a inclusão de textos gerados por modelos de linguagem em longa escala, como o *ChatGPT*, a menos que o texto produzido seja apresentado como parte de uma análise experimental (*Top AI conference bans ChatGPT in paper submissions*, 2023). Seguiu a mesma linha de entendimento de que tais ferramentas seriam perigosas, ao menos neste momento, na medida em que não se poderia garantir a integridade do texto resultante. A Itália, *v. g.*, proibiu o uso do *ChatGPT* em virtude da proteção de dados dos italianos (Barbosa, 2023)⁸, bem como o fez a *Springer Nature*, que é a principal editora de revistas científicas do mundo (Capozzi, 2023)⁹.

Por outro lado, Luciano Rossoni (2022), Editor da Revista Eletrônica de Ciência Administrativa – RECADM, analisa o fenômeno também sobre outro aspecto, destacando como a ferramenta pode ser valiosa para “liberar os cientistas da tarefa árdua e nem sempre central da escrita”. Isto é, para o autor, ferramentas com o *ChatGPT* permitem que os cientistas e escritores possam economizar tempo e esforço para redigirem resumos, artigos e outros documentos científicos, podendo utilizá-los no processo criativo e na geração de novas ideias. Sem contar que, para Rossoni (2022), a escrita automatizada poderia ser menos suscetível a erros e incongruências tipicamente humanas.

8 A notícia pode ser conferida em: <https://forbes.com.br/forbes-tech/2023/03/o-que-levou-a-italia-a-banir-o-chatgpt/?amp>. (Accessed: 29 March 2023).

9 Esta informação é retirada de: <https://olhardigital.com.br/2023/01/29/pro-maior-editora-de-revistas-cientificas-proibe-chatgpt-como-coautor/amp/>. (Accessed: 29 March 2023).

Independentemente do caminho que essas discussões tomarem, não há dúvidas de que a Inteligência Artificial Generativa, especialmente com *ChatGPT*, vem influenciando na cultura e no mercado, sendo apontada como o futuro da pesquisa (Stokel-Walker, 2023). Porém, engana-se quem pensa que se trata de uma possibilidade futura; uma vez que a ferramenta já está à disposição, sendo sua utilização e aprimoramento inexoráveis. Inclusive no âmbito do Direito Processual, em que periodicamente são descobertas novas utilizações, que assustam ou animam, encantem ou decepcionam.

3. A experiência brasileira com a utilização da Inteligência Artificial no Processo Judicial

Se a IA Generativa pode ser considerada uma novidade no processo judicial, a experiência brasileira na utilização da IA já vem se consolidando há alguns anos. Passou-se o tempo que a Inteligência Artificial era utilizada no âmbito do processo apenas para pesquisa de jurisprudência. Especialmente com o impulso do fluxo de dados advindo da digitalização da justiça¹⁰, o processo, de uns anos para cá, passou a possibilitar a inclusão de ferramentas e sistemas inteligentes desde a fase pré-processual até a fase executiva.

A IA Forte, que possui propósito geral e ultrapassa as tarefas específicas e padronizadas, adquirindo aos poucos as mais variadas competências humanas, ainda não está sistematicamente incorporada no processo judicial, muito embora existam alguns projetos direcionadas a esse objetivo¹¹. Um exemplo importante a ser mencionado é a utilização da IA na seara probatória, em que já é possível deparar-se com robôs que são desenvolvidos para o auxílio na reconstrução de fatos e para conceder prognósticos através de probabilidades (Gaio Junior and Silva, 2023). Os exemplos citados por Antônio Pe-

10 Importante frisar que IA não se confunde com digitalização da justiça e com processo eletrônico, sendo conceitos distintos. Ou seja, uma justiça digitalizada e a utilização do processo eletrônico não significam, necessariamente, a utilização de Inteligência Artificial, muito embora sejam seus impulsionadores.

11 Sobre o tema consultar Ribeiro, 2022, p. 230 e ss.

reira Gaio Júnior e Fábria Antonio Silva (2023) são os Robôs Stevie, Eco e Pierce-IGTT, Data Mining e Alibi.

Já a IA Fraca atua em áreas restritas e especializadas, já com resultados importantes, está presente, por exemplo, nos Programas “Victor” (2018) e “Rafa” (2022), desenvolvidos pelo Supremo Tribunal Federal, com atuação em âmbito recursal. Geralmente, a Inteligência Artificial Fraca está associada às atividades externas à prestação jurisdicional, aquelas que não envolvem necessariamente uma tomada de decisão jurisdicional.

Outro exemplo de IA Fraca pode ser verificado na Defensoria Pública do Estado do Rio de Janeiro, que vem testando programa denominado “Caixa de Intimação com Inteligência Artificial (CICIAR)”, o qual busca uma solução para um dos gargalos da jurisdição (Parreira et al, 2022). Essa ferramenta foi criada para auxiliar defensores públicos no tratamento e priorização de atendimento das intimações que são recebidas no Programa *Verde*, o qual a Defensoria Pública daquele Estado utiliza para gerenciar o acompanhamento dos seus processos. De acordo com Parreira et al (2022), “o CICIAR tem como diferencial categorizar as intimações recebidas sem acessar seu conteúdo, o que iniciaria a contagem do prazo para o seu atendimento”, auxiliando no planejamento do defensor.

Mais modelos de IA utilizados no processo judicial também foram apresentados por John O. McGinnis e Russell G. Pearce, em revolucionário texto de 2014. Os professores da Northwestern University School of Law e da Fordham University School of Law, respectivamente, descrevem cinco áreas em que a IA poderá ser utilizada no processo judicial: (I) descoberta de documentos e dados (*discovery*), (II) pesquisa jurídica (*legal search*); (III) redação de documentos (*document generation*); (IV) elaboração de relatórios (*brief generation*); e previsão do resultado em casos futuros (*prediction of case outcomes*) (McGinnis and Pearce, 2014).

Especificamente no Poder Judiciário, Daniel Henrique Arruda Boeing e Alexandre Moraes da Rosa (2020, pp. 95-100), distinguem três grupos de usos diferentes de IA. Os “robôs-

-classificadores” possuem a função de encontrar materiais úteis para que os seres humanos possam redigir os pronunciamentos, como as decisões. Por sua vez, os “robôs-relatores” extraem e condensam informações relevantes de documentos, podendo catalogar peças processuais, diferenciando o que é texto legal, o que é jurisprudência e o que é descrição de fatos, além de outras tarefas de igual importância. Por fim, os “robôs-julgadores”, produzem resultados que serão considerados a própria decisão judicial.

Esse último grupo de IA mencionado por Boeing e Rosa, corresponde, efetivamente, a IA Generativa. As possibilidades, em diversos âmbitos, de modelos de IA como o ChatGPT já foram indicadas acima, evidenciando a profunda interação com o usuário por meio de perguntas e respostas, abarcando temas diversos e complexos. Indo além, para poder responder à pergunta formulada neste artigo, parte-se a entender como vem ocorrendo a utilização da IA Generativa, especialmente do ChatGPT, no mundo jurídico.

4. A utilização da IA Generativa no mundo jurídico

O Direito também está com os olhos abertos à IA generativa, especialmente ao ChatGPT. A ferramenta ganhou destaque com feitos como ter passado na prova da OAB brasileira e alcançado consistentes resultados na prova necessária a tornar-se Advogados nos Estados Unidos. Nesta, após experimento em que o Chatbot alcançou êxito em quatro testes para ingressos em faculdades de Direito na *University of Minnesota* (Kelly, 2023), foi testado no *Bar Exam*, atingindo desempenho bastante similar a média dos graduados que realizam a prova (Merritt and Bilek, 2023).

Na experiência brasileira, o experimento foi pensado e conduzido por Daniel Marques, filósofo, advogado e presidente da Associação Brasileira de Lawtechs e Legaltechs – AB2L. Na prática, Marques enviou as perguntas da prova do ano de 2022, de múltipla escolha, para o ChatGPT, pedindo que escolhesse entre uma das alternativas e justificasse o

motivo da escolha. Conforme relatou o presidente da AB2L, o sistema teria acertado 48 das 80 perguntas da prova, sendo considerado aprovado na primeira fase do exame (Migalhas, 2023).

Nos escritórios de Advocacia, esse tipo de ferramenta pode e já vem sendo utilizado em algumas frentes, para além da pesquisa e da redação de documentos. É o caso, a título exemplificativo, da firma Allen & Overy, com sede em Londres, que realizou um experimento em setembro de 2022 com o sistema de IA Generativa da *OpenAI*. Conforme David Wakeling, chefe do grupo de inovação do escritório, alguns advogados da Allen & Overy passaram a utilizar a ferramenta para responder a simples perguntas sobre legislação, redigir documentos e escrever mensagens a clientes (Stokel-Walker, 2023).

Apesar do experimento iniciar com alguns advogados apenas, a ferramenta tornou-se uma sensação, de modo que, em fevereiro de 2023, cerca de 3.500 funcionários da Allen & Overy, em 43 escritórios, a estavam utilizando, com um total de 40.000 consultas. Para Wakeling, o que mostram esses números iniciais é o início de uma mudança paradigmática, em que a IA Generativa é uma realidade no mundo jurídico por ser a ele muito adequada (Stokel-Walker, 2023).

No âmbito do Poder Judiciário, a IA Generativa também já foi testada. Após a divulgação pública do ChatGPT, uma das primeiras notícias da sua utilização em sede de decisão judicial foi na Colômbia, em que o juiz Juan Manuel Padilla, titular do “Juzgado 1 Laboral del Circuito de Cartagena” utilizou da ferramenta para redigir uma decisão judicial.

Com o ChatGPT atuando na criação de argumentos que sustentassem sua fundamentação, a sentença proferida foi favorável a uma cidadã que ajuizara a demanda contra “Salud Total EPS”, buscando a exoneração do pagamento de taxas, terapias e transporte ao hospital pelo seu filho, alegando que não possuía recursos suficientes (Quevedo, 2023). A utilização da ferramenta, segundo a magistrada colombiana, estaria juridicamente apoiada nos seguintes argumentos que constaram no próprio julgado:

El propósito de incluir estos textos producto de la IA no es en manera alguna reemplazar la decisión del Juez. Lo que realmente buscamos es optimizar los tiempos empleados en redacción de sentencias, previa corroboración de la información suministrada por IA”, destaca la sentencia antes de exponer las consultas que concretamente le planteó a la aplicación (Quevedo, 2023).

Ou seja, o que amparou a utilização do ChatGPT naquela Corte colombiana foi “a otimização do tempo dos servidores na redação das sentenças” (Quevedo, 2023). Assim como no Brasil, está claro que a utilização da IA Generativa na seara de decisões certamente vem (e virá) respaldada por um discurso de maior celeridade na prestação da jurisdição.

O que chama a atenção neste programa não é só a capacidade da Inteligência Artificial para ingressar no mundo jurídico, uma vez que sistemas com tecnologias semelhantes já atuam em atividades jurisdicionais, desde leitura e separação de documentos até a redação de modelos de decisão. E nesses âmbitos vem atuando bastante bem, tanto que o número de projetos de IA desenvolvidos e em desenvolvimento no Poder Judiciário brasileiro é de 111, distribuídos em 53 tribunais, conforme dados coletados pelo CNJ no programa do Programa Justiça 4.0 (CNJ, 2022).

O que impressiona é a profundidade, a riqueza de detalhes das respostas e a coesão textual. Certamente, nesses aspectos, é possível dizer que supera parte significativa dos juristas de carne e osso, especialmente por basear-se em dados (sejam eles verdadeiros ou não) e não em meras opiniões. E com essa afirmação não se pretende desmerecer o trabalho do jurista humano, mas traz a reflexão acerca da antiga discussão acerca da substituição das profissões jurídicas, o que não cabe nas linhas desse artigo.

Possibilidades como as acima mencionadas levam à necessidade de repensar os institutos processuais, para atualizá-los ao momento atual, em que a utilização da Inteligência Artificial já é uma realidade latente. Ainda mais quando se trata da IA Generativa, com todas as suas possibilidades revolucionárias (algumas acima mencionadas). Por isso, de nada adianta tentar impedir esse avanço tecnológico na seara

jurídica; até porque parece ser uma realidade sem volta, que demanda uma ressignificação¹² e não desprezo e paralisação.

5. ChatGPT e a fundamentação das decisões judiciais

A utilização da Inteligência Artificial Generativa no âmbito do processo judicial introduz o debate acerca de diversos institutos processuais, dos quais este artigo destaca a fundamentação das decisões judiciais. Esse destaque mostra-se necessário diante da pertinência teórica e prática, sendo que é a decisão judicial que possibilita o exercício da jurisdição. Afinal, se pudesse o magistrado valer-se de “fundamentos” criados (ou compilados) pelo ChatGPT para decidir, qual é o impacto disso na indispensável motivação? Será que se estaria pondo fim à fundamentação das decisões judiciais?

Essa discussão faz lembrar a obra do autor espanhol Jordi Nieva Fenoll (2018), uma das primeiras sobre o tema, intitulada “Inteligencia Artificial y Proceso Judicial”, lançada pela Marcial Pons em 2018. Naquela obra precursora, Fenoll dialoga sobre a possibilidade de alcançarmos um patamar de desaparecimento da fundamentação das decisões judiciais. Não seria surpresa que isso ocorresse porque, para o autor,

se aceptará sin más la resolución que proponga la herramienta de inteligencia artificial, que se tendrá por suficientemente justificada si el algoritmo fue configurado correctamente y nadie lo discute (Fenoll, 2018, p. 117).

Quer dizer, o que defende Fenoll é que a Inteligência Artificial poderia produzir resultados tão espetaculares em termos de construção de textos – como de decisões judiciais – que a mágica de justificar residiria na programação do sistema e não no resultado final obtido. A evidência dessa possibilidade residiria nos casos repetitivos, em que se poderia prescindir da motivação em detrimento de modelos preesta-

12 O termo “ressignificação” é aqui utilizado no sentido de significar um repensar dos conceitos tradicionais do Direito Processual, buscando adequá-los ao momento tecnológico atual. O tema da ressignificação dos institutos processuais a luz da Inteligência Artificial é trabalhado em Fröhlich (2023).

belecidos e sem a devida reflexão em cada situação pontual, que pouco diverge da construção modelar (Fenoll, 2018, pp. 117-118). Nas palavras do professor espanhol:

Y el final de todo ese camino es la desaparición de la motivación en estos casos, sustituida por el buen funcionamiento que se suponga del algoritmo en escala general. De hecho, hoy en día ya se podría haber prescindido de la motivación en estos casos reiterativos, que son además bastante numerosos, lo que, por cierto, hubiera ahorrado bastante trabajo burocrático al personal del tribunal (Fenoll, 2018, p. 118).

Essa perspectiva, de certa forma, ganha estofo com as ferramentas de IA Generativa, as quais efetivamente criam textos que poderão ser utilizados em decisões judiciais. Porém esquece ela que a fundamentação das decisões judiciais não pode ser comparada a um julgamento de valor sobre o bom ou o mau funcionamento do algoritmo. A conclusão que adviria desta análise estaria restrita à explicação e a transparência a respeito dos passos lógico-matemáticos percorridos pela máquina para decidir (algoritmo). A fundamentação, por outro lado, demanda interpretação e conexões que não residem apenas na programação.

Especificamente ao que propõe este artigo, resume-se a discussão em perguntar se bastaria ao magistrado utilizar-se da ferramenta para que estivesse fundamentada a decisão e fosse cumprido o requisito previsto no artigo 93, inciso IX, da Constituição Federal¹³ e nos artigos 11¹⁴ e 489, caput, inciso II¹⁵, do atual Código de Processo Civil. Em um primeiro momento, a resposta parece ser negativa.

Ocorre que o modelo de fundamentação das decisões incorporado pelo Código de Processo Civil possui uma estru-

13 O dispositivo da Constituição Federal do Brasil (art., 93, IX), prevê que “todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões, [...]”. (Brasil, 1988).

14 O artigo 11 do Código de Processo Civil brasileiro, reproduzindo a previsão constitucional, dispõe que “Todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões, sob pena de nulidade” (Brasil, 2015).

15 A referida previsão (art. 489, caput, II, do CPC) diz ser elemento essencial da sentença “os fundamentos, em que o juiz analisará as questões de fato e de direito” (Brasil, 2015).

tura fechada, muito embora seu conteúdo seja aberto. A forma que deve ser proferida e fundamentada uma decisão judicial está disposta na lei processual civil, no artigo 489, estando na previsão legal (e na sua efetiva observância) um dos principais limites que garante a legitimidade do que for decidido.

Nesse sentido, antes de avançar, cabe aqui retornar uma definição basilar do conceito de fundamentação, que diferencia sua importância endo e extraprocessual. Se, por um lado (função endoprocessual), é a fundamentação que garante e expõe a observância, pelo magistrado, das fases processuais e das garantias fundamentais caras ao Estado Democrático de Direito; permitindo impugnar por meio de recurso; por outro (função extraprocessual), possui a fundamentação importante valor político, que permite à sociedade conhecer e analisar a forma pela qual o poder jurisdicional está sendo exercido, inclusive para servir de modelo de comportamento jurídico aos operadores do direito, indicando, ao lado da lei, qual seria o comportamento desejado perante o ordenamento jurídico. Sem contar que se trata a fundamentação de cláusula pétrea, por encontrar abrigo no artigo 60, §4º, da Constituição Federal de 1988, que veda alterar a Constituição no intuito de macular na separação dos Poderes (Silva, 2007).

Por isso é que fundamentação das decisões não pode ser deixada de lado nem mesmo em um ambiente de desenvolvimento e utilização da Inteligência Artificial Generativa. Quer dizer, mesmo que utilizada a ferramenta, não há um passe livre, um locus de ausência de fundamentação. Deverão, sim (e talvez ainda mais), serem observados a Constituição e o Código de Processo Civil ao dispor aquela soberanamente sobre a obrigatoriedade de fundamentar e esta especificamente acerca dos critérios para considerar uma decisão fundamentada.

Nesse patamar ingressam os requisitos negativos previstos no primeiro parágrafo do já mencionado artigo 489 da lei processual civil. É a não incidência em nenhuma das hipóteses em que a decisão é considerada “não-fundamentada” que garante uma decisão concreta, estruturada e completa, nas belas palavras de Marinoni, Arenhart e Mitidiero, (2022, p. 610). Por isso, a utilização de ferramentas de IA Generati-

va (mesmo que como auxiliares), jamais poderá deixar de observar os incisos do art. 489, parágrafo primeiro, do Código de Processo Civil.

A primeira limitação à decisão judicial reside no primeiro inciso do parágrafo primeiro do artigo 489 do CPC, o qual não considera fundamentada a decisão judicial que “se limitar à indicação, à reprodução ou à paráfrase de ato normativo, sem explicar sua relação com a causa ou a questão decidida” (Brasil, 2015). Bastante semelhante a essa disposição, o inciso V do artigo 489 do CPC dá o adjetivo de não fundamentada à decisão que “se limitar a invocar precedente ou enunciado de súmula, sem identificar seus fundamentos determinantes nem demonstrar que o caso sob julgamento se ajusta àqueles fundamentos” (Brasil, 2015). Ambas as previsões referem-se a um modelo lógico-dedutivo de aplicação da lei ou do precedente ao caso concreto, o que é insuficiente no atual estágio do Estado Democrático de Direito brasileiro.

Certo é que as ferramentas de IA Generativa muito superaram os seres humanos na habilidade de confrontar premissas e alcançar, com base nelas, decisões, o que é característico do modelo lógico-dedutivo, ainda mais quando se exige uma decisão provinda de um juiz, ou sistema, que seja imparcial¹⁶. Caso fosse esse o modelo adotado para a fundamentação da decisão judicial, não haveria qualquer ressalva a ser feita ao ChatGPT, plenamente capaz de receber um questionamento e, por meio da sua programação, analisar os dados à disposição, ou inputs (leis e precedente) e, a partir daí gerar seu output (resultado ou decisão). Sobre isso, preocupa-se Luís Greco, dizendo que “A ideia do juiz-robô seria uma reedição do sonho racionalista do *juiz como máquina de subsunção* – um sonho do qual já acordamos há mais de duzentos anos” (Greco, 2020, p. 43).

Vale aqui a reflexão segundo a qual já estamos, há bastante tempo, trabalhando com juízes-robôs, basta ver os números trazidos na obra *Justiça em Números* do CNJ, do ano de 2022 (CNJ, 2022), para percebermos que nossos magistrados, infelizmente, estão obrigados a julgar uma enormidade de pro-

16 Sobre as estreitas relações havidas entre fundamentação e imparcialidade, consultar Ribeiro, 1998, pp. 19-22.

cesso, *v. g.*, o Tribunal de Justiça de São Paulo julgou em 2020 a extraordinária marca de 4.352.829 processos (TJSP, 2021), enquanto a 8ª Câmara Cível do Tribunal de Justiça do Rio Grande Sul, que é de Família, julgou a incrível quantidade de 1.000 processo em uma sessão (TJRS, 2020). Nossa preocupação, portanto, é com os robôs-Juízes e não com os juízes-robôs.

Contudo, a simples comparação entre premissas para a indicação de um texto legal, de um precedente ou de uma súmula, sem a devida relação esmiuçada com o caso concreto, não é suficiente para cumprir o requisito do artigo 489 da lei processual civil. Especialmente, não cumpre os mencionados incisos I e V do parágrafo primeiro do dispositivo.

Ou seja, ao utilizar-se do ChatGPT, não basta ao juiz resumir em algumas palavras o que compreendeu do caso e introduzir no sistema para que seja indicado o texto legal ou o precedente que possa incidir. O resultado não será considerado fundamentação, uma vez que, como dito, decidir não é somente aplicar um texto legal ou mesmo comparar premissas. Para que haja, de fato, fundamentação, a norma jurídica deve ser relacionada ao caso concreto, indicando as circunstâncias fáticas em jogo. É isso, certamente, o mais difícil para o *Chat-GPT* que precisará entrar em contato com todos os dados (fatos e provas, por exemplo) existentes nos autos, realizando escolhas dentro dos limites hermenêuticos.

Ademais, também não pode o sistema de IA Generativa decidir com o emprego de conceitos jurídicos indeterminados, sem explicar o motivo concreto de sua incidência no caso. Da mesma forma, não poderá invocar motivos que se prestariam a justificar qualquer outra decisão. Essas duas práticas encontrariam, respectivamente, vedação nos incisos II e III do parágrafo primeiro do artigo 489 do Código de Processo Civil, considerando-se a decisão que não observar essas limitações nula por ausência de fundamentação.

Ao primeiro (inciso II do artigo 489, parágrafo primeiro, do CPC) chama-se atenção o fato de que a expressão do resultado alcançado pela IA Generativa deve ser não só em linguagem natural, como, e principalmente, em linguagem técnico-jurídica. É dizer, não pode ser uma resposta genérica

ou mesmo com gírias e abreviações, a qual é geralmente utilizada na grande parte dos *sites* disponíveis na internet e que servem de banco de dados para o modelo. Ou seja, a conclusão alcançada não pode se desgarrar do seu encaixe jurídico, que lhe garante legitimidade, levando em consideração o Direito vigente.

É o que mesmo que se afirma com relação ao inciso III do parágrafo primeiro do artigo 489 do CPC. Muito embora a matéria prima do resultado encontrado pelo *ChatGPT* esteja na junção e correspondência entre dados encontrados na rede mundial de computadores, só poderão ser aproveitados aqueles que efetivamente se relacionam ao caso e as suas peculiares condições. Sem contar que devem ser utilizados apenas aqueles dados de fontes confiáveis, limitação que não se sabe ao certo se é (e como é) observada pelo *ChatGPT*.

Francisco José García-Peñalvo (2023) sustenta que o *ChatGPT*, assim como já ocorrera com outras ferramentas disponíveis na *internet* – como a Wikipédia, por exemplo – pode se tornar obsoleto na medida em que terá acesso a quaisquer fontes de informação abertas que estiverem disponível na internet. Isso é, para o autor, um problema, na medida em que poderia utilizar materiais redigidos por qualquer pessoa, sem análise crítica do que foi selecionado e sem a devida atribuição às fontes originais (García-Peñalvo, 2023). Causa preocupação, nestas condições, a utilização dessas fontes genéricas e sem análise crítica, em fundamentações de decisões judiciais.

Indo além, ferramentas de IA Generativa, em regra, irão analisar todas as informações disponíveis em seus bancos de dados para responder ao questionamento que lhes foram propostas. Desta forma, quanto melhor a base de dados, melhor será a fundamentação fornecida pelo sistema inteligente. Sem contar que também levarão em consideração o que aprenderam com os feedbacks (respostas) que já receberam. Isso quer dizer que, teoricamente, não deixará de seguir (ao menos de observar) enunciado de súmula, jurisprudência ou precedente invocado pela parte, desde que foram alimentadas com essas informações. Utilizar-se da IA Generativa para auxílio na

decisão não violaria, ao que parece, à primeira parte do inciso VI do §1º do artigo 489 do CPC.

Por outro lado, dois problemas quanto a isso podem ser destacados. O primeiro é se, de fato, informações sobre enunciado de súmula, jurisprudência ou precedente invocado pela parte possuem o mesmo peso no sistema. E, o outro, é com relação a segunda parte do inciso VI do parágrafo primeiro, do artigo 489 do CPC, isto é, se o *ChatGPT* será capaz de demonstrar a existência de distinção no caso em julgamento ou a superação do entendimento.

Por fim, quanto ao problema de não enfrentar todos os argumentos deduzidos no processo capazes de, em tese, infirmar a conclusão adotada pelo julgador (artigo 489, §1º, inciso IV, do CPC), a resistência que se teve ao cumprimento dos demais incisos pelas ferramentas de IA Generativa parece um pouco menor. Em tese, todas as perguntas colocadas no sistema serão enfrentadas pelo *ChatGPT* (da forma que forma, como visto acima), desde que sejam adicionadas ao sistema, não havendo o perigo de “esquecer-se” de analisar um argumento, como seguidamente percebemos quando a decisão nasce de um juiz humano. Até porque, repita-se: em regra, todos os *inputs* (dados) colocados nos sistemas serão analisados para alcançar-se o resultado, no que a ferramenta se diferencia do julgador humano.

Sem contar um cenário provável em que os “Embargos de Declaração” serão drasticamente reduzidos ou eliminados, posto que dificilmente haverá omissão, bem como a dificuldade de recorrer-se de uma decisão proferida pela IA Generativa. Poderia ela, inclusive, ser comparada ao Juiz Hércules de Dworkin. Porém, isto é tema para um próximo artigo.

Isto quer dizer, a despeito da ironia, que a IA Generativa assume erros quando confrontada, aprendendo a cada novo feedback que recebe. Com relação ao julgador humano, a atuação diária nos fóruns brasileiros permite perceber que o artigo 489, §1º, do CPC é reiteradamente esquecido. Será que, com a utilização do *ChatGPT*, poderá ser esse maltratado artigo ganhar novos contornos?

6. Considerações finais (e uma experiência)

O que se pretendeu neste artigo foi analisar como a fundamentação das decisões judiciais está sendo e pode ser impactada pela Inteligência Artificial Generativa, cujo principal exemplo é o do aclamado *ChatGPT*. Trata-se de tema complexo, ainda mais difícil para juristas, que cada vez mais são obrigados a encarar novas possibilidades de Inteligência Artificial. Luís Greco (2020, p. 17), sobre isso, questiona ironicamente se

faz sentido que um jurista, cuja experiência com tecnologia se reduz ao manuseio de programas de edição de texto e de navegação na rede mundial de computadores, pergunte sobre a possibilidade fática de construir um juiz robô.

O presente artigo permite, de certa forma, traçar alguns pressupostos para o início de resposta à pergunta de Greco. Sobre a “possibilidade fática”, o atual cenário tecnológico foi traçado pelos apontamentos acerca do estado da arte da Inteligência Artificial e do desenvolvimento de IA generativa. Porém, este artigo foi além e enfrentou algumas das (im)possibilidades jurídicas. Quanto a estas, adentrou-se no tema da fundamentação das decisões judiciais, especialmente sob uma perspectiva de critérios negativos, analisando o tema com amparo nas delimitações do artigo 489, §1º, do CPC. Afinal, traz esse dispositivo as hipóteses em que uma decisão (utilizando-se IA ou não) não poderá ser considerada fundamentada.

O que já estamos percebendo é que a diferença entre ciência humana e exata está, rapidamente, desaparecendo. O campo de incidência de cada uma delas está cada vez mais tênue. Será que a jurisdição apresentada pela Constituição Federal é a prometida pelos algoritmos? É esta a ideia de justiça apregoada pela Constituição Federal? Uma justiça algorítmica? Qual seria, então, a diferença entre justiça humana e justiça algorítmica? Não existem, ainda, respostas para estas perguntas.

O que se pode concluir, muito embora trate-se de uma pesquisa que por sua essência e evolução, jamais será exauriente, foi que a utilização unicamente do *ChatGPT* para decidir, neste momento, não atende ao dever de fundamentar. Como

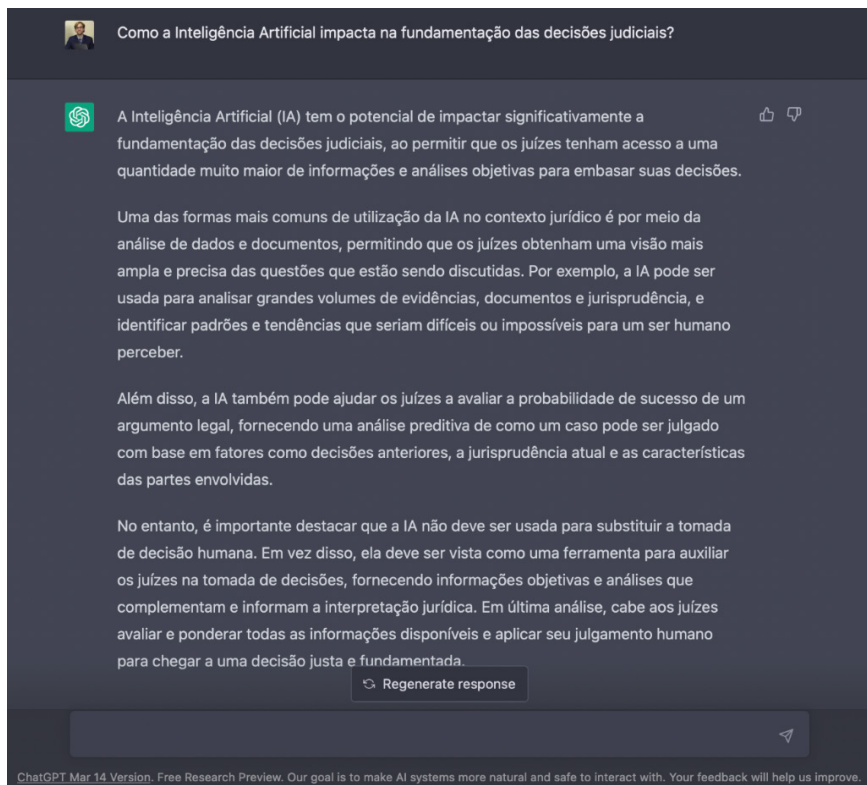
Luciano Rossoni (2023) bem considerou ao finalizar o editorial da *Revista Eletrônica de Ciência Administrativa – RECADM*, no qual utilizou-se da referida ferramenta para redigir o texto, “todo o texto foi produzido pela inteligência artificial, mas ele se mostrou extremamente dependente das ideias selecionadas por mim e das minhas orientações e construções de conjunções textuais”. E arrematou “Isso mostra como a inteligência artificial pode ser uma ferramenta poderosa, mas ainda precisa ser orientada e supervisionada pelo ser humano” (Rossoni, 2022).

Quanto à fundamentação das decisões, apesar de alguns testes iniciais com a utilização da IA Generativa, ainda se trata de incumbência humana, o que não afasta a sua utilização em uma série de tarefas de auxílio, como uma ferramenta poderosa à disposição do julgador. Neste caso, diferentemente do que defende Nieva Fenoll (2018), a fundamentação estará ainda mais fortalecida, atendendo ao preceito constitucional e, principalmente, observando o artigo 489 do CPC e os incisos do seu parágrafo primeiro.

Para finalizar, a título ilustrativo e para sanar curiosidades, vai colacionada abaixo uma singela experiência proposta pelos autores deste artigo: a resposta do próprio *ChatGPT* (2023), versão *free* e não *plus* (que certamente aprimoraria a solução apresentada) à pergunta-problema que norteou esse trabalho¹⁷. Com efeito, o que surpreendeu é que algumas conclusões deste trabalho coincidem com o que disse a IA, especialmente quando conclui a ferramenta que “ela [a I] deve ser vista como uma ferramenta para auxiliar os juízes na tomada de decisões, fornecendo informações objetivas e análises que complementam e informam a interpretação jurídica”. Apesar disso, segue o questionamento que norteou a escolha do tema deste artigo: o que fundamenta a afirmação do *ChatGPT*?

17 Para evitar qualquer discussão sobre autoria, esclareça-se que os autores deste artigo não utilizaram da resposta da IA para redação do presente texto, tendo sido a pergunta ao *ChatGPT* formulada, a título de experiência e por curiosidade acadêmica, posteriormente à redação do texto final.

Figura 1 – Experiência dos autores: resposta do *ChatGPT* (2023) à pergunta-problema que norteou esse trabalho.



Fonte: ChatGPT, 2023.

Referências

- About* [2023?] OpenAI [online]. Available at: <https://openai.com/about> (Accessed: 18 March 2023).
- Aras, V. (2020) 'A Inteligência Artificial e o direito de ser julgado por humanos' in Pinto, H. A.; Guedes, J. C. and César, J. P. C. *Inteligência artificial aplicada ao processo de tomada de decisões*. Belo Horizonte; São Paulo: D' Plácido, 2020, pp. 85-130.

- Barbosa, A. (2023) ‘O que levou a Itália a proibir o ChatGPT?’, *Forbes* [online]. Available at: <https://forbes.com.br/forbes-tech/2023/03/o-que-levou-a-italia-a-banir-o-chat-gpt/?amp> (Accessed: 29 March 2023).
- Boeing, D. H. A. and Rosa, A. M. (2020) *Ensinando um robô a julgar: pragmática, discricionariedade, heurísticas e vieses no uso de aprendizado de máquina no Judiciário*. Florianópolis: Emais Academia.
- Brasil (1988) *Constituição da República Federativa do Brasil de 1988*. Brasília, DF: Presidência da República [online]. Available at: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm (Accessed: 27 March 2023).
- Brasil (2015) *Lei n. 13.105, de 16 de março de 2015: Código de Processo Civil*. Brasília, DF: Presidência da República [online]. Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm (Accessed: 27 March 2023).
- Capozzi, B. (2023) ‘Maior editora de revistas científicas proíbe ChatGPT como coautor’, *Olhar Digital* [online]. Available at: <https://olhardigital.com.br/2023/01/29/pro/maior-editora-de-revistas-cientificas-proibe-chatgpt-como-coautor/amp/> (Accessed: 29 March 2023).
- ChatGPT* (2023?) OpenAI [online]. Available at: <https://chat.openai.com/chat>. Acesso em 18 mar. 2023.
- Christian, B. (2011) *The most Human Human*. New York: Doubleday.
- Conheça a Rafa – Ferramenta de IA para classificar ações conforme ODS da Agenda 2030* (2022), *Youtube*, Brasília, 16 May 2022. 1 vídeo (2 h 1 min, 23 s) [online]. Available at: <https://www.youtube.com/watch?v=Fhl0snROqfU&t=214s>. (Accessed: 18 March 2023).
- Conselho Nacional de Justiça (2022). *Justiça em Números 2022 (ano-base 2021)*, *CNJ* [online]. Available at: <https://www.cnj.jus.br/wp-content/uploads/2022/09/justica-em-numeros-2022-1.pdf> (Accessed: 30 March 2023).
- Fenoll, J. N. (2018) *Inteligencia Artificial y proceso judicial*. Madrid: Marcial Pons.

Fröhlich, A. V. K. (2023) *Fundamentação das Decisões Judiciais e Inteligência Artificial: uma resignificação ao Direito Processual atual e futuro*. Londrina: Editora Thoth.

Gaio Júnior, A. P. and Silva, F. A. (2023) 'Direito, Processo e Inteligência Artificial': diálogos necessários ao exercício da jurisdição. *Revista Eletrônica de Direito Processual – REDP*. Periódico Quadrimestral da Pós-Graduação Stricto Sensu em Direito Processual da UERJ, Rio de Janeiro, v. 24, n 1, ano 17, Janeiro-abril de 2023, p. 60-99 [online]. Available at: <https://www.e-publicacoes.uerj.br/index.php/redp/article/view/72240/44601> (Accessed: 19 March 2023).

García-Peñalvo, F. J. (2023) 'La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico'. *Education In The Knowledge Society (Eks)*, Salamanca, Ediciones Universidad de Salamanca, 24, p. 1-9, 6 February 2023. [online]. Available at: <https://revistas.usal.es/tres/index.php/eks/article/view/31279/29185> (Accessed: 28 March 2023).

Greco, L. (2020) *Poder de julgar sem responsabilidade de julgador: a impossibilidade jurídica do juiz-robô*. São Paulo, SP: Marcial Pons.

Introducing ChatGPT (2023) OpenAI [online]. Available at: <https://openai.com/blog/chatgpt> (Accessed: 18 March 2023).

Jesus, C. (2016) 'AI Lawyer "Ross" Has Been Hired By Its First Official Law Firm', *Futurism*, 5 November 2016 [online]. Available at: <https://futurism.com/artificially-intelligent-lawyer-ross-hired-first-official-law-firm> (Accessed: 18 March 2023).

Kelly, S. (2023) 'ChatGPT passes exams from law and business schools', *CNN Business*, 26 January 2023 [online]. Available at: <https://edition.cnn.com/2023/01/26/tech/chatgpt-passes-exams/index.html> (Accessed: 25 March 2023).

Latorre, J. I. (2019) *Ética para máquinas*. Barcelona: Editora Planeta.

- Lubowitz, J. H. (2023) 'ChatGPT, An Artificial Intelligence Chatbot, Is Impacting Medical Literature', *Arthroscopy: the Journal of Arthroscopy and Related Surgery*, 38, 31 January 2023, pp. 1-2 [online]. Available at: [https://www.arthroscopyjournal.org/article/S0749-8063\(23\)00033-6/fulltext](https://www.arthroscopyjournal.org/article/S0749-8063(23)00033-6/fulltext) (Accessed: 25 March 2023).
- Mantegna, M. (2020) *Inteligencia ARTEficial: Creatividad computacional, inteligencia artificial generativa y derechos de autor*. Tesis de la Maestría en Propiedad Intelectual e Innovación, Universidad de San Andrés, Argentina, Novembro de 2020 [online]. Available at: <https://repositorio.udea.edu.ar/jspui/handle/10908/17998> (Accessed: 23 March 2023).
- Marinoni, L. G., Arenhart, S. C. and Mitidiero, D. (2022) *Código de Processo Civil Comentado*, 8th edn. revised, updated and expanded. São Paulo: Thomson Reuters Brasil.
- Marrit, D. J. and Bilek, M. L. (2023) 'ChatGPT Almost Passed the Bar, But Competent Lawyers Do Much More', *Blomberg Law*, 13 February 2023 [online]. Available at: <https://news.bloomberglaw.com/business-and-practice/chatgpt-almost-passed-the-bar-but-competent-lawyers-do-much-more> (Accessed: 25 March 2023).
- McGinnis, J. and Pearce, R. G. (2014). 'The Great Disruption: How Machine Intelligence Will Transform the Role of Lawyers in the Delivery of Legal Services'. *Fordham Law Review*, v. 3041, *Northwestern Public Law Research Paper*, n. 14-17, *Fordham Law Legal Studies Research Paper* n. 339387 [online]. Available at: <https://ssrn.com/abstract=2436937> (Accessed: 11 September 2023).
- Melo, J. O. (2016) 'Escritório de advocacia estreia primeiro "robô-advogado" nos EUA', *Consultor Jurídico*, 16 May 2016 [online]. Available at: <https://www.conjur.com.br/2016-mai-16/escritorio-advocacia-estreia-primeiro-robo-advogado-eua> (Accessed: 18 March 2023).

Migalhas (2023) ‘Advogado virtual? ChatGPT consegue “aprovação” na primeira fase da OAB’, *Redação Migalhas*, 22 February 2023 [online]. Available at: <https://www.migalhas.com.br/quentes/381875/advogado-virtual-chatgpt-consegue-aprovacao-na-primeira-fase-da-oab> (Acesso 25 March 2023).

Minsky, M. (1968) *Semantic information processing*. Cambridge: MIT Press.

Parreiras, M., Vasconcellos, A., Mangeli, E., Yamamoto, E., Xexéo, G., Metello, I. Costas, L., Marques, P. and Souza, J. (2023) ‘Inteligência artificial aplicada para o aumento da produtividade no atendimento de intimações’. *Anais do X Workshop de Computação Aplicada em Governo Eletrônico*. Niterói: Sociedade Brasileira de Computação [online]. Available at: <https://sol.sbc.org.br/index.php/wcge/article/view/20721/20548> (Accessed: 20 March 2023).

Pérez-Lunõ, A. E. (2020) ‘Inteligencia artificial y posthumanismo’ in Bravo, Á. S. (coord.), *Derecho, inteligencia artificial y nuevos entornos digitales*. Sevilla: AADMDS, p. 18.

Quevedo, G. (2023) ‘La sentencia de Colombia y su incursión en el derecho’, *LexLatin*, 8 February 2023 [online]. Available at: <https://lexlatin.com/noticias/chatgpt-en-el-derecho-sentencia-de-colombia> (Accessed: 18 March 2023).

Resultados Pesquisa IA no Poder Judiciário – 2022 (2022) in CONSELHO Nacional de Justiça: Justiça 4.0. [online]. Available at: https://paineisanalytics.cnj.jus.br/single/?appid=9e4f18ac-e253-4893-8ca1-b81d8af59ff6&sheet=b8267e5a-1f1f-41a7-90ff-d7a2f4ed34ea&lang=pt-BR&theme=IA_PJ&opt=ctxmenu,currsel&select=language,BR. (Accessed: 25 March 2023).

Ribeiro, D. G. (1998) *Provas Atípicas*. Porto Alegre: Livraria do Advogado.

Ribeiro, D. (2020) *O novo processo civil: presente e futuro*. Londrina, PR: Thoth.

Ribeiro, D. (2022) *O novo processo civil: presente e futuro*. Londrina, PR: Thoth. p. 230 e ss.

- Rio Grande do Sul. Tribunal de Justiça do Estado do Rio Grande do Sul (2020). *8ª Câmara Cível realiza sessão virtual* in Comunicação [online]. Available at: <https://www.tjrs.jus.br/novo/noticia/8a-camara-civel-realiza-sessao-virtual/> (Accessed em: 30 March 2023).
- Rossoni, L. (2022) 'A inteligência artificial e eu: escrevendo o editorial juntamente com o ChatGPT', *Revista Eletrônica da Ciência Administrativa – RECADM*, Curitiba, 21(3), pp. 399-405, September-December. Disponível em <http://periodicosibepes.org.br/index.php/recadm/article/download/3761/1167> (Accessed: 23 March 2023).
- Russel, S. and Norvig, P. (2013). *Inteligência artificial*. Rio de Janeiro: Elsevier Editora Ltda. *E-book* [Online].
- Salvagno, M., Taccone, F. S. and Gerli, A. G. (2023) 'Can artificial intelligence help for scientific writing?', *Critical care*, 27(1), pp. 1-5.
- São Paulo, Tribunal de Justiça do Estado de São Paulo (2021). *TJSP fecha 2020 com mais de 4,3 milhões de processos julgados in Notícias* [online]. Available at: <https://www.tjsp.jus.br/Noticias/Noticia?codigoNoticia=63248> (Accessed: 30 March 2023).
- Silva, B. O. (2007) *A garantia fundamental à motivação da decisão judicial*. Salvador: JusPodivm.
- Stokel-Walker, C. (2023) 'Generative AI Is Coming for the Lawyers: large law firms are using a tool made by OpenAI to research and write legal documents. What could go wrong?', *Wired*, 21 February 2023 [online]. Available at: <https://www.wired.co.uk/article/generative-ai-is-coming-for-the-lawyers> (Accessed: 26 March 2023).
- Thorp, H. H. 'ChatGPT is fun, but not an author, Science, Washington D. C.', 379(6630), 26 January 2023, p. 313 [online]. Available at: <https://www.science.org/doi/full/10.1126/science.adg7879> (Accessed: 23 March 2023).

Top AI conference bans ChatGPT in paper submissions (and why it matters), *Venture Beat*, 5 January 2023 [online]. Available at: <https://venturebeat.com/ai/thats-so-meta-ml-conference-debates-use-of-chatgpt-in-papers/> (Accessed: 23 March 2023).

Turing, A. M. (1950) 'Computing machinery and intelligence' *Mind: A quarterly review of psychology and philosophy*, New Series, Oxford, 59(236), October, pp. 433-460 [online]. Available at: https://www.jstor.org/stable/2251299?origin=JSTOR-pdf&seq=1#page_scan_tab_contents (Accessed: 30 July 2021).

INTELIGENCIA ARTIFICIAL Y TEORÍA DE LA EVOLUCIÓN. REFLEXIONES EN TORNO A LA VIABILIDAD DE LA *GLOBAL LEX DIGITALIS*

Carlos de Cores Helguera¹

DOI: 10.29327/5312711.1-8

Resumen: El artículo plantea la cuestión de la inteligencia artificial en clave de la teoría de la evolución de Charles Darwin, presentando primero el desarrollo de Yuvel Noah Harari, y proponiendo a continuación una visión diferente, que en lugar de pronosticar el sometimiento del homo sapiens a la inteligencia artificial, sostiene que ella seguirá estando sometida al control humano, lo que reafirma la necesidad de normas y principios ético jurídicos que se sintetizan en la idea de una “*global lex digitalis*”.

Palabras clave: Inteligencia Artificial. Teoría de la evolución. Charles Darwin.

Sumario: 1. Introducción. 2. Necesidad de orientaciones deontológicas para las regulaciones de la Inteligencia Artificial. 3. La crisis de los límites de la Inteligencia Artificial. 4. Primera Tesis de Harari: los seres humanos no somos más que algoritmos. 5. Segunda Tesis de Harari: los algoritmos de la Inteligencia Artificial evolucionarán hacia el paso de homo sapiens a homo Deus. 6. Tercera Tesis de Harari: será el fin del humanismo, ya que las nuevas tecnologías despojarán a los humanos de su autoridad y la trasladarán a algoritmos no humanos. 7. La evolución de la especie humana hacia el dominio de la Inteligencia Artificial. 8. La evolución no es incompatible con la libertad humana. 9. La inteligencia artificial dominará al homo sapiens, o éste dominará a la Inteligencia Artificial?

1 Doctor en Derecho por la Universidad Pontificia Comillas, Madrid. Director de Derecho Privado de la Universidad Católica del Uruguay.

10. Conclusión: de Homo sapiens a Homo Deus, o de Homo Sapiens a Deus? Referencias.

1. Introducción

Como enseña Díaz Alabart (2018, p. 14), la inteligencia artificial (más conocida por sus siglas, IA), es una ciencia que trata de estudiar el comportamiento inteligente a través de modelos matemáticos. Según la autora, no es conveniente identificar la inteligencia de una artificial con la de los humanos; no son comparables; funcionan de forma muy diferente: las máquinas hoy por hoy, están totalmente especializadas, y solo desarrollan comportamientos muy específicos; los humanos, en cambio, tienen una inteligencia que podríamos llamar generalista en cuanto que permite desarrollar comportamientos en todos los aspectos de la vida; tienen sentido común y emociones. Por ello, no es factible traducir al lenguaje de los algoritmos los principios éticos de las personas (Díaz Alabart, 2018, p. 14).

La distinción propuesta por Díaz Alabart coincide básicamente con la diferencia entre Inteligencia Artificial débil y fuerte. La propia de las máquinas es la Inteligencia Artificial débil: la máquina está programada para desarrollar distintas modalidades pero de carácter repetitivo. En cambio el hombre tiene la aptitud de adaptarse y generar enorme cantidad de lecturas y sentidos. El término inteligencia proviene del latín “*inte legere*”, que significa “leer entre”, o sea, aportar sentidos.

La cuestión que ocupa a la academia refiere a la Inteligencia Artificial fuerte, o sea, la que refiere a la capacidad de las máquinas de aportar y desarrollar sentidos. Ello se ha logrado a partir de la construcción de máquinas computacionales siguen la matriz de las redes neuronales artificiales² siguiendo la inspiración que provee el comportamiento del cerebro humano (las neuronas y sus conexiones), creando modelos artificiales que solucionan problemas complejos, mediante técnicas algorítmicas convencionales.

2 Ver Redes Neurales. Universidad de Salamanca. Available at: <http://avellano.fis.usal.es/~lalonso/RNA/index.htm>.

Las redes neuronales artificiales actuales son susceptibles de “entrenamiento”, mediante el cual se proporciona a la red neuronal modelos de entrada (equivalentes a las preguntas o problemas) y los correspondientes modelos de salida (las respuestas apropiadas). Tras una gran cantidad de eventos de entrenamiento, la red aprende asociar los modelos de entrada a los de salida, adquiriendo la capacidad de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Las máquinas preparadas para ello pueden aprender a partir de sus experiencias, sacar conclusiones y actuar en consecuencia. Para ello se utilizan algoritmos de aprendizaje o entrenamiento, los cuales, al igual que en el aprendizaje humano, mejoran con el tiempo, en base a una mayor “experiencia”, y la ventaja es que son completamente autónomos.³

Hay diferentes tipos de algoritmos de aprendizaje. El bot, es un término que procede de la palabra “robot”, y que puede definirse como un programa informático que, de forma automática, efectúa tareas repetitivas a través de Internet, cuya realización por una persona física sería imposible o muy tediosa. El chatbot es un tipo de bot (bot conversacional) capaz de mantener un diálogo mediante un lenguaje natural con un humano (o con otro chatbot) y que ha sido diseñado y programado para poder interpretar el motivo o la intención de la conversación. Por ello, también puede entender las respuestas del destinatario y, a su vez, responderle correctamente.⁴

Es común que hagamos preguntas a chatbots en nuestros dispositivos incluso celulares.

3 Díaz Alabart (2018) cita como un buen ejemplo práctico de este aprendizaje automático el caso del desafío a cinco partidas del robot AlphaGo contra el 14 veces campeón mundial Lee Sidol. En las primeras tres partidas el humano que había jugado perdió rápidamente contra la máquina, pero para la cuarta partida (después de reunirse en consulta con otros maestros en el juego) preparó una estrategia totalmente diferente, iniciar una serie de jugadas que podríamos denominar «extrañas» o incluso «absurdas» de acuerdo con lo habitual, frente a las que la máquina no supo reaccionar, perdiendo la partida. Pero en la quinta y última partida el robot ya había podido aprender con esa nueva experiencia, y el humano, sin tener ya a su favor el “factor sorpresa” la perdió.

4 Para los conceptos de bot y chatbot v. <https://planetachatbot.com/que-es-un-chatbot>.

2. Necesidad de orientaciones deontológicas para las regulaciones de la Inteligencia Artificial

El progreso incesante de la Inteligencia Artificial en su versión fuerte y la acumulación de aprendizaje por parte de las redes neuronales artificiales, ha llevado a la pregunta sobre sus límites. Desde hace mucho se habla de las reglas de Asimov, también llamadas reglas de la robótica, la primera y principal de las cuales da cuenta de que la Inteligencia Artificial está sometida al ser humano y no puede dañar ni permitir que un ser humano sufra daño.

Es hoy día un lugar común que la enorme expansión de la Inteligencia Artificial hace necesario avanzar en el consenso de principios éticos que orienten las regulaciones jurídicas que dicten los distintos estados, o las organizaciones internacionales que puedan aparecer en el marco de lo que Felipe Barcarollo (2019, p. 227) llamó *global lex digitalis*. En su tesis doctoral, Barcarollo dedica un párrafo entero a esta cuestión.⁵

Así como a partir de la fragmentación del Imperio Romano, los comerciantes que habitaban sus territorios fueron desarrollando la llamada *lex mercatoria*, a partir de la fragmentación del mundo global, es plausible que se vaya generando un conjunto de reglas jurídicas y normas éticas que Barcarollo llama *global lex digitalis*, un nuevo paradigma regulatorio que debe tener como centro el ser humano y su permanente capacidad de analizar las repercusiones de la Inteligencia Artificial (Barcarollo, 2019, p. 265).

Citando a Gunther Teubner (2016), Barcarollo (2019, p. 111) sostiene que estas normas regularían los límites de la intervención de la técnica en la vida humana, haciendo que la máquina o artificio sea un facilitador o aliado del ser humano para la mejor convivencia en la sociedad, postulando la necesidad de que estos límites consideren los impactos éticos, legales y sociales (Ethical, Legal and Social Impacts – ELSI) de la Inteligencia Artificial (Barcarollo, 2019, p. 211). De este modo,

5 4.3 Dimensões Éticas e Regulatórias em Inteligência Artificial: desafios ao Direito na contemporaneidade, preparando as estruturas de uma *global lex digitalis*. (Barcarollo, 2019, p. 217).

en el futuro previsible, la Inteligencia Artificial permanecería subordinada a la inteligencia humana natural.

3. La crisis de los límites de la Inteligencia Artificial

Ahora bien: esta visión, que gira en torno al concepto de que la Inteligencia Artificial está sometida a los límites ético-jurídicos impuesto por el hombre, e encuentra confrontada con la que se contiene en las obras de Yuvel Noah Harari. Este autor ha escrito una trilogía compuesta por tres libros: *De animales a dioses*, *Homo Deus* y *21 lecciones para el siglo XXI*, de los que se han impreso millones de ejemplares en muchas lenguas. Se trata de una investigación realmente impresionante, que acumula una enorme cantidad de información que va llevando al autor a realizar una lectura de la dirección en que se orienta la Historia de la Humanidad.

Las principales tesis que – en mi lectura – sostiene Harari (2017, p. 377), relativas a la inteligencia artificial, son las siguientes:

1. Los seres humanos son algoritmos.
2. Los algoritmos propios de la Inteligencia Artificial evolucionarán perfeccionándose hasta el punto de que dominarán a los humanos.
3. Será el fin del humanismo, ya que las nuevas tecnologías despojarán a los humanos de su autoridad y la trasladarán a algoritmos no humanos (Harari, 2017, p. 377).

De acuerdo con esta visión, la Inteligencia Artificial evolucionará hacia un estado en el cual no estará contenida por límite alguno, sino que por el contrario, dominará al hombre, al homo sapiens. Ello implica una contradicción con el escenario imaginado para la *global lex digitalis*.

Admitiendo que el trabajo de Harari se fundamenta en una importante investigación, analizaremos una por una sus tesis, proponiendo en cada caso una breve reflexión crítica, con la intención de generar un insumo para una discusión que puede ser útil para la configuración de las regulaciones

ético-jurídicas de la Inteligencia Artificial, esto es, en el proceso de cristalización de la *global lex digitalis*.

4. Primera Tesis de Harari: los seres humanos no somos más que algoritmos

Harari sostiene claramente que

durante los últimos años los biólogos han llegado a la firme conclusión de que el hombre [...] es también un algoritmo. [...] Los algoritmos que controlan a los seres humanos operan mediante sensaciones, percepciones y pensamientos. Y exactamente el mismo tipo de algoritmos controla a cerdos, babuinos, nutrias y gallinas (Harari, 2017, p. 101).

Agrega que

[...]el 99 % de nuestras decisiones (entre ellas, las elecciones más importantes de la vida, relacionadas con los cónyuges, carreras y hábitats) las toman los refinadísimos algoritmos que llamamos sensaciones, emociones y deseos (Harari, 2017, p. 103).

Evidentemente, la ciencia ha ido clarificando los procesos electro-bio-químicos que están en la base de los comportamientos humanos. Sin embargo, descartar que los seres humanos estemos dotados de libre arbitrio, dependiendo totalmente de nuestra constitución material, no parece una conclusión necesaria. Nos permitimos señalar que ello puede representar la opinión del autor, pero observamos una falacia de no seguimiento. El descubrimiento de los funcionamientos de nuestros procesos vitales no es incompatible lógicamente con la existencia del libre arbitrio. La propia afirmación de que nuestras decisiones las toman algoritmos en “el 99 % de los casos”, sin un desarrollo que justifique por qué ello ocurre exactamente en el 99 % de los casos, por una parte arroja dudas sobre lo que ocurriría en el 1 % restante, y por otra muestra el sesgo que condiciona la tesis del autor, ya que no aparece ningún fundamento para esta determinación cuantitativa.

Hemos realizado el ejercicio de plantear a la propia Inteligencia Artificial, en este caso, un chatbot, el ChatGPT, las cuestiones que propone Harari.

Frente a la pregunta: “El hombre puede ser considerado un conjunto de algoritmos?”, el ChatGPT ha respondido:

En cierto sentido, se puede decir que el hombre es un conjunto de algoritmos. Los algoritmos son secuencias de instrucciones precisas y ordenadas que se utilizan para realizar una tarea específica. En el cerebro humano, también hay secuencias de instrucciones precisas y ordenadas que se utilizan para realizar diferentes tareas, desde procesar información sensorial hasta tomar decisiones y controlar el movimiento. En este sentido, se podría decir que el cerebro humano utiliza algoritmos para procesar información y realizar acciones. Sin embargo, también es importante tener en cuenta que el cerebro humano es mucho más complejo que cualquier algoritmo que haya sido diseñado hasta ahora. El cerebro tiene una capacidad de adaptación, creatividad y pensamiento crítico que aún no se ha logrado replicar en una máquina. Por lo tanto, aunque se pueda decir que el hombre utiliza algoritmos en cierto sentido, no se puede reducir la complejidad y la singularidad de la experiencia humana a un simple conjunto de algoritmos.⁶

La propia Inteligencia Artificial en su estado actual, parece no admitir la conclusión absoluta de Harari.

A continuación preguntamos al ChatGPT: “Somos sujetos libres o nuestros actos están determinados por factores que no controlamos?” La respuesta fue:

Esta es una pregunta que ha sido debatida por filósofos, psicólogos y científicos durante siglos. La respuesta depende en gran medida de la perspectiva filosófica que se adopte. Algunas teorías sostienen que nuestros actos están determinados por factores que están más allá de nuestro control, como la genética, el ambiente en el que crecemos y las experiencias que hemos tenido a lo largo de nuestra vida. Según esta perspectiva, nuestras acciones y decisiones están predestinadas por estos factores, y por lo tanto no somos sujetos libres. Por otro lado, otras teorías sostienen que sí tenemos libre albedrío y que podemos tomar decisiones conscientes y racionales que no están completamente determinadas por

6 Consulta realizada el 14.4.2023 hora 11.51.

factores externos. Según esta perspectiva, aunque estamos influenciados por nuestra genética y nuestro entorno, todavía tenemos la capacidad de elegir cómo reaccionar y cómo tomar decisiones.

En última instancia, la respuesta a esta pregunta es compleja y depende de muchas variables. Lo que es seguro es que nuestros actos son el resultado de una interacción compleja entre muchos factores, algunos de los cuales están dentro de nuestro control y otros que no lo están. Aunque es posible que no tengamos control total sobre todas las variables que influyen en nuestras decisiones, todavía podemos tomar decisiones conscientes y hacer elecciones éticas y responsables”.⁷

De modo que no es tan claro que, en el futuro cercano o previsible, deje de ser relevante la cuestión ética, la dimensión de la responsabilidad humana, por lo que nos inclinamos a descartar la tesis de que no será útil la reflexión sobre la *global lex digitalis*.

5. Segunda Tesis de Harari: los algoritmos de la Inteligencia Artificial evolucionarán hacia el paso de homo sapiens a homo Deus

Sostiene Harari (2017, pp. 56-58) que la Inteligencia Artificial provocará el ascenso de humanos a dioses por medio de cualquiera de estos tres caminos: ingeniería biológica, ingeniería ciborg e ingeniería de seres no orgánicos. Así, los ingenieros biológicos reescribirán nuestro código genético. La ingeniería ciborg fusionará el cuerpo orgánico con dispositivos no orgánicos. Y en un enfoque más audaz, podría pronosticarse que el homo sapiens prescindirá totalmente de las partes orgánicas y producirá seres inteligentes totalmente inorgánicos (Harari, 2017, pp. 56-58).

Ello conduce a Harari a formularse estas preguntas:

¿Si los programas informáticos alcanzan una inteligencia sobrehumana y unos poderes sin precedentes: deberemos empezar a valorar esos programas más de lo que valoramos a los humanos? ¿Será aceptable, por ejemplo, que una inteligencia artificial explote a los seres humanos e incluso los mate para

7 Consulta realizada el 14.4.2023 Hora 11.16.

favorecer sus propias necesidades y deseos? [...] ¿qué probabilidad hay de que entidades no humanas lleguen alguna vez a rivalizar con nosotros y superarnos? (Harari, 2017, p. 117).

En respuesta a estos interrogantes, concluye Harari que

cuando la tecnología nos permita remodelar la mente humana, Homo sapiens desaparecerá, la historia humana llegará a su fin y se iniciará un tipo de proceso completamente nuevo, que [...] no podemos ni imaginar. [...] En el siglo XXI, el tercer gran proyecto de la humanidad será adquirir poderes divinos de creación y destrucción, y promover el Homo sapiens a Homo Deus (Harari, 2017, p. 59).

Se comprende fácilmente cómo la visión del autor se encontraría en contradicción con la primera ley de Asimov, quitando toda viabilidad o utilidad a la labor de construcción de una *global lex digitalis*.

6. Tercera Tesis de Harari: será el fin del humanismo, ya que las nuevas tecnologías despojarán a los humanos de su autoridad y la trasladarán a algoritmos no humanos

Según Harari (2017, p. 81), el concepto “humanismo” tiene un significado preciso, que entendemos restrictivo y por tanto equivocado. Para Harari (2017, p. 81), el humanismo es aquella cosmovisión “que sacraliza la vida, la felicidad y el poder del Homo sapiens”, en virtud de lo cual

l intento de conseguir la inmortalidad, la dicha y la divinidad no hace más que llevar los antiguos ideales humanistas a su conclusión lógica. Sitúa abiertamente sobre la mesa lo que durante mucho tiempo hemos mantenido oculto bajo la servilleta. [...] El humanismo (el culto a la humanidad) ha conquistado el mundo. Pero el auge del humanismo contiene asimismo las semillas de su caída. Mientras que el intento de mejorar a los humanos hasta convertirlos en dioses lleva al humanismo a su conclusión lógica, deja al descubierto simultáneamente sus defectos inherentes [...] (Harari, 2017, p. 81).

Agrega Harari (2017, p. 115) que mientras que la revolución agrícola dio origen a religiones teístas, la revolución científica dio origen a religiones humanistas, en las que los humanos sustituyeron a los dioses.

No parece ajustarse a la verdad que el humanismo pueda ser definido como la filosofía, la teoría y la práctica de convertir a los seres humanos en dioses. Este predicado no puede ser cierto, por ejemplo, para ninguno de los humanismos religiosos.

Harari (2017, p. 277) menciona al humanismo liberal, al humanismo socialista y al humanismo nazi. Esta presentación de las clases principales en que se dividiría el concepto de humanismo sería en mi concepto demostrativa por sí sola de que se está aludiendo únicamente a una pésima caricatura de lo que puede considerarse humanismo.

Por ejemplo: el humanismo propio de toda la tradición religiosa judío cristiana, se basa en la relación entre el hombre (y su colectivo, el pueblo) y Dios, por lo que no puede sostenerse que tienda a la confusión entre el hombre y Dios, y menos que lo coloque como finalidad del hombre convertirse en Dios. Precisamente, en la imagen que se observa en el relato bíblico, la circunstancia de que el hombre desee “ser como Dios” o convertirse en dios, constituye un apartamiento de la relación originalmente establecida.

7. La evolución de la especie humana hacia el dominio de la Inteligencia Artificial

Las ideas de Darwin tienen un enorme peso en la concepción de Harari. Al fin de su tomo *Homo Deus*, al formalizar las conclusiones de sus extendidos estudios, Harari menciona a dos autores: Darwin y Turing. Darwin aporta la idea de la evolución, Turing aporta la idea de la inteligencia artificial computacional que constituirá el núcleo de la nueva etapa progresiva que culminará con la evolución de la especie humana.

La respuesta monoteísta tradicional es que solo los sapiens poseen un alma eterna. ... sin embargo – sostiene Harari

(2017, p. 119) – los últimos descubrimientos científicos contradicen de plano este mito monoteísta. ...

Harari expresa:

Darwin nos ha privado de nuestra alma. Si uno entiende de verdad la teoría de la evolución, comprende que el alma no existe. Se trata de una idea terrible, no solo para los cristianos y musulmanes devotos, sino también para los muchos laicos que no tienen ningún dogma religioso claro, pero que no obstante quieren creer que cada humano posee una esencia individual eterna que permanece inalterada a lo largo de la vida y que puede sobrevivir intacta a la muerte. [...]

Lamentablemente, la teoría de la evolución rechaza la idea de que mi yo verdadero sea una esencia indivisible, inmutable y potencialmente eterna.

[...]

(La evolución se basa en cambios genéticos pequeños. Por lo tanto) la biología no puede explicar el nacimiento de un bebé con un alma eterna hijo de unos padres que no tuvieran una pizca de alma.

De ahí que la existencia de almas no pueda armonizar con la teoría de la evolución. ... Esto aterra a un gran número de personas que prefieren rechazar la teoría de la evolución antes que renunciar a su alma (Harari, 2017, pp. 122-123).

Sin embargo, no está para nada claro que en la concepción de Darwin, esté exenta la idea del alma o la de una inteligencia creadora. Esta posibilidad parece inaceptable para Harari. Éste comenta una encuesta de Gallup en la que explica que una parte importante de los estudiantes universitarios cree que «los humanos pudieron haber evolucionado a partir de formas de vida previas en un proceso que durase millones de años, pero que Dios orquestó todo el espectáculo (Harari, 2017, p. 120).

Parece claro que Harari prefiere la hipótesis que excluye un ente creador. No obstante, al inicio de su obra, Darwin (1859, página inicial) trae una cita de Bacon que desmiente esta visión de las cosas: “Todo hombre debe progresar y ser experto (proficient) tanto en el libro de la palabra de Dios, como en el libro de las obras de Dios”. (Bacon: *Advancement of Learning*).

Por otra parte, al final, el propio Darwin realiza una conclusión que intenta compatibilizar la teoría de la evolución, con la apertura a la posibilidad de un Creador inteligente:

A mi modo de ver, concuerda mejor con lo que sabemos de las leyes impresas en la materia por el Creador, que la producción y extinción de los habitantes pasados y presentes del mundo se debieron a causas secundarias, como las que determinan el nacimiento y la muerte de el individuo. Cuando veo a todos los seres no como creaciones especiales, sino como descendientes directos de unos pocos seres que vivieron mucho antes de que se depositara el primer lecho del sistema Silúrico, me parece que se ennoblecen.

...

Y como la selección natural obra únicamente por y para el bien de cada ser, todas las dotes corporales y mentales tenderán a progresar hacia la perfección (Darwin, 1859, final).

8. La evolución no es incompatible con la libertad humana

Para demostrar, o al menos, cuestionar, que la teoría de la evolución de las especies no es incompatible con la posibilidad del alma y de Dios, es suficiente aludir a los escritos de Pierre Teilhard de Chardin, que no es mencionado por Harari.

Pero se trata de una cuestión con siglos de acumulación. Ya en el Renacimiento, el debate sobre la interacción entre la intervención de Dios en la Historia y la libertad humana, se formalizó en la denominada *polémica de auxiliis*, llevada a cabo en los albores de la Edad Moderna, que concluyó en la proposición de la diferencia entre causas primeras y causas segundas. Que las causas segundas estén regidas por una rígida determinación susceptible de ser comprendida y pronosticada por la ciencia, no significa que quede resuelta la cuestión de la causa primera, o sea, de que haya existido o no un Dios creador.

Una obra que resulta significativa para entender la centralidad de la idea de libertad en la relación del Hombre con Dios, es la *Concordia liberi arbitrii cum gratiae donis*, de Luis de Molina S. J. (1535-1600), que contiene una amplia reflexión sobre el problema de la relación entre la libertad del hombre y la gracia de Dios. Esta obra propone una compleja forma de

explicar la conciliación entre la potestad divina y su conocimiento de los hechos futuros, y la libertad del hombre.

En la obra de Luis de Molina la cuestión de la libertad viene a ser considerada esencial a la misma definición del ser humano. Dios, como causa primera, crea el Universo según un Plan y todo lo creado se organiza conforme a ese Plan, destinado, en definitiva, al Bien, según la Providencia Divina. Pero una vez creado, el Universo se rige por sus propias leyes; no sigue dependiendo permanentemente de la intervención divina arbitraria, como ocurría en la mitología griega.

La tesis teológica central en la *Concordia liberi arbitrii* es que Dios, como causa primera, crea el mundo y el Hombre; luego las causas segundas se desarrollan sin intervención directa de Dios, aunque todo tiende a cumplir el Bien según la Divina Providencia.

Esta explicación teológica conduce a la exaltación de la libertad humana, ya que la relación del hombre con Dios no puede ser efectiva sin la aceptación libre por parte del hombre. En este contexto, Dios crea al hombre y lo dota de libertad. Para que pueda afirmarse que el hombre es realmente libre, es preciso que el Creador se autolimite en cuanto a la posibilidad de actuar en la cadena de causas segundas, porque si Dios estuviera permanentemente actuando para alterar los vaivenes de la Historia, el hombre no sería verdaderamente libre ni protagonista.

En la teología cristiana, la Encarnación de Dios y su inserción en Historia humana, que constituye el centro de la experiencia cristiana, no se realiza en un contexto de conquista y avasallamiento divinos, sino de una paradójal manifestación (epifanía) de humildad y debilidad. Esta idea central es planteada por San Pablo en su epístola a los filipenses (Capítulo 2 versículos 6 a 8) en los siguientes términos: “Cristo, que era de condición divina, no consideró esta igualdad con Dios como algo que debía guardar celosamente: al contrario, se anonadó a sí mismo, tomando la condición de servidor y haciéndose semejante a los hombres. Y presentándose con aspecto humano, se humilló hasta aceptar por obediencia la muerte y muerte de cruz”.

La clásica figura del nacimiento de Jesús en Belén da cuenta de un Dios hecho hombre que no impone su mensaje por la fuerza, sino que realiza un llamado sutil cuyo seguimiento depende de la libre decisión del hombre. De lo que resulta un planteo doctrinal que da un relieve propio y muy importante a la iniciativa y a la libertad humana; o, dicho de otro modo, se trata de un planteo doctrinal desde la libertad de la persona.

Ello implica que, en esta cosmovisión, los procesos humanos (Grice-Hutchinson, 1952, p. 46) adquieren una cierta autonomía. Así también el Hombre, creado por Dios como causa primera, queda constituido en agente libre.

El discurso laicista radical sigue por el camino de negar la existencia de Dios y la de aquel acto creador de la esfera de la causa primera, oponiendo la fe en Dios, a la Ciencia. Las proposiciones de Harari parecerían adscribirse a esta perspectiva, desde la escéptica expresión ya mencionada “Dios orquestó todo”, es decir, Dios, causa primera, orquestó la secuencia de las causas segundas.

Sin embargo, tanto la Reforma protestante como la Contrarreforma católica coinciden en que en ningún caso puede admitirse una determinación de la conducta humana por parte de Dios, ya que ello sería incompatible con la libertad del Hombre. La intervención de Dios es sugerencia, invitación, inspiración. El Mesías de Dios no forzará nunca al hombre, sino que respetará siempre su libertad: “no quebrará la caña cascada, ni apagará el pabilo humeante” (Biblia, Isaías, 42.3).

Ello que confiere su lugar propio a la Ciencia. Las causas segundas no son manejadas o “orquestadas” por Dios, lo que determina la plena aplicación de las leyes científicas, ya que de otro modo, el hombre no sería un sujeto libre.

En consecuencia, la actitud religiosa del cristiano creyente, tanto protestante como católico, por más científico que sea, no sólo no es incompatible con la afirmación más radical de la libertad humana, sino que, por el contrario, ella constituye su axioma más característico.

9. La inteligencia artificial dominará al homo sapiens, o este dominará a la Inteligencia Artificial?

El ChatGPT parece entender que el homo sapiens, al menos en el horizonte previsible, seguirá mandando.

Ante la pregunta: ¿Qué peligros implica la Inteligencia Artificial?, se enumeran peligros: sesgo y discriminación, desempleo, cyberseguridad, fallos y errores, uso malintencionado. Y se agrega:

Para abordar estos peligros, es importante que se desarrollen políticas y regulaciones claras en torno a la IA y que se fomente la colaboración entre los expertos en tecnología, la industria y los responsables de la toma de decisiones.⁸

La conclusión es, pues, que la Inteligencia Artificial requiere políticas y regulaciones externas, otorgando sentido a la tarea de proponer una *global lex digitalis* que oriente dichas regulaciones en forma compatible con los valores y los derechos humanos.

Estas políticas y regulaciones no pueden ser elaboradas por la propia Inteligencia Artificial.

Según el ChatGPT:

Aunque la Inteligencia Artificial (IA) puede ser utilizada para analizar grandes cantidades de datos y ayudar en la toma de decisiones, todavía no es posible que la IA elabore por sí sola políticas y regulaciones en torno a su propio uso. La IA es una herramienta diseñada por seres humanos para ser utilizada por ellos, y aunque puede ser muy útil en la toma de decisiones, todavía no tiene la capacidad de tener una comprensión completa y profunda de los valores y objetivos humanos, así como del contexto social, político y cultural en el que se encuentra.

Es importante que los seres humanos sigan desempeñando un papel activo en la toma de decisiones en torno al desarrollo y uso de la IA, y que se establezcan mecanismos adecuados para supervisar y regular su uso de manera efectiva. Esto puede incluir la creación de grupos de expertos, la consulta de expertos en ética y derechos humanos, y la colaboración entre los gobiernos, la industria y la sociedad civil para asegurar que el desarrollo de la IA esté guiado por principios éticos y humanos.⁹

8 Consulta ChatGPT Hora 11.08.

9 Consulta ChatGPT Hora 11.09.

Intentando profundizar en la materia, la pregunta final es: Podrá en el futuro la IA tener una comprensión completa y profunda de los valores y objetivos humanos, así como del contexto social y político en el que se encuentra?

Es difícil predecir el futuro de la Inteligencia Artificial (IA) y si alguna vez será posible que tenga una comprensión completa y profunda de los valores y objetivos humanos, así como del contexto social y político en el que se encuentra. La IA ha avanzado rápidamente en las últimas décadas, y ha logrado grandes avances en áreas como el procesamiento del lenguaje natural, la visión por computadora y el aprendizaje automático.

Sin embargo, es importante tener en cuenta que la comprensión de los valores y objetivos humanos implica una comprensión profunda de la cultura, la historia y las interacciones sociales humanas, así como de los valores y creencias personales. Es poco probable que la IA pueda adquirir esta comprensión de manera autónoma, ya que se basa en la experiencia y la reflexión humana.

Por lo tanto, es probable que la comprensión de los valores y objetivos humanos siga siendo una tarea principalmente humana en el futuro previsible, aunque la IA puede seguir siendo una herramienta útil para ayudar en la toma de decisiones informadas y éticas. Será importante seguir discutiendo y reflexionando sobre los límites y las capacidades de la IA, y asegurarse de que su desarrollo y uso estén guiados por principios éticos y humanos.¹⁰

De estos diálogos, es posible proponer, al menos como enunciado susceptible de generar una discusión seria, que – en el futuro previsible – siga teniendo sentido el esfuerzo de generar consensos – en los ámbitos propios de la Inteligencia Artificial, que exceden las fronteras nacionales, un conjunto de orientaciones éticas y jurídicas que pudieran ser integradas en un cuerpo normativo como el llamado *global lex digitalis*.

10. Conclusión: de *Homo sapiens* a *Homo Deus*, o de *Homo Sapiens* a *Deus*?

Realmente la evolución conducirá a la desaparición del homo sapiens? Naturalmente, no es serio sostener que la so-

10 Consulta ChatGPT Hora Id. Hora 11.12

lución a esta pregunta quede resuelta mediante el recurso a la propia Inteligencia Artificial que presuntamente superará al hombre hasta el punto de determinar la extinción de la especie humana y su sustitución por nuevas entidades dominadas por la Inteligencia Artificial.

No es responsabilidad del homo sapiens observar que la Inteligencia Artificial es un producto humano y puede debe ser orientado según lo estime conveniente el ser humano?

Aludiendo a la contratación automatizada, si bien en la superficie la decisión robótica o algorítmica de contratar aparece *prima facie* desligada de toda voluntad humana, e imputable directamente a una máquina, inmediatamente se cae en la cuenta de que siempre son una inteligencia y una voluntad humanas las que predisponen los medios algorítmicos (Helgera and Damini, 2020). Ello hace plausible que el debate en torno a la *global lex digitalis* pueda continuar y desarrollarse en el futuro próximo.

Referências

- Barcarollo, F. (2019) *Inteligência Artificial e a Gramática Ético-Jurídica da Sociedade (Pós)-Humana*. Tese de Doutorado em Direito, Programa de Pós-Graduação em Direito da Universidade do Vale do Rio dos Sinos – Unisinos.
- Bíblia, N. T. (2023) 'Isaias 4,3', in *Bíblia Sagrada*. Translated by Fernando. 3rd edn. São Paulo: Editora NVI.
- Darwin, C. (1859) *On the Origin of Species*. London: John Murray.
- Helgera, C. C. and De Cores Damini, C. C. (2020) 'Derecho privado patrimonial e inteligencia artificial' in Miranda, J. E., Hupffer, H. M. and Engelmann, W. (eds.). *Direito e Inteligência Artificial: O Desafio Ético No Emprego Das Novas Tecnologias*. Curitiba: Brazil Publishing, pp. 41-68.
- Diaz Alabart, S. (2018) *Robots y responsabilidad civil*. Madrid: Reus.

Grice-Hutchinson, M. (1952) *The School of Salamanca, Readings in spanish monetary theory, 1544-1605*, Oxford: Clarendon Press.

Harari, Y. N. (2017) *Homo Deus: Breve historia del mañana*. Barcelona: Editorial Debate.

Teubner, G. (2016) *Fragmentos constitucionais: constitucionalismo social na globalização*. São Paulo: Saraiva.

Universidade de Salamanca. *Redes Neurales* [online]. Available at: <http://avellano.fis.usal.es/~lalonso/RNA/index.htm>. (Accessed: 14 August 2023).

CONSECUENCIAS JURÍDICAS DEL USO DE CHATGPT EN LA ACCIÓN DE TUTELA EN COLOMBIA¹

Sebastián Díaz Bolívar²

Dany Steven Gómez Agudelo³

DOI: 10.29327/5312711.1-9

Resumen: Este texto tiene por objetivo establecer los efectos legales de utilizar ChatGPT para revolver una acción de tutela en Colombia. La metodología utilizada consistió en seleccionar la técnica de la revisión documental e indagar información en bases de datos como science direct, Google académico, vlex y Scopus. Se concluye que ChatGPT no tiene el alcance suficiente para resolver cuestiones judiciales como la acción de tutela, sin embargo, se puede utilizar como una herramienta complementaria en la labor del funcionario judicial, además es importante que el Consejo Superior de la Judicatura, brinde mayor capacitación a todos los jueces en relación con el uso eficaz y responsable la tecnología, con el fin de garantizar los derechos fundamentales de las personas ante el uso de la inteligencia artificial.

- 1 Este capítulo se deriva de la investigación denominada “efectivización de las garantías constitucionales en la implementación de la justicia digital, durante la pandemia del Covid-19 en Colombia”, la cual fue financiada por la Universidad Católica Luis Amigó, sede Medellín, Colombia, en el año 2023.
- 2 Magíster en Derecho Procesal Contemporáneo. Doctorando en Derecho Procesal Contemporáneo de la Universidad de Medellín. Coordinador del área de derecho privado civil II del Programa de Derecho de la Universidad Católica Luis Amigó. Miembro del grupo de investigación *Jurídicas y Sociales*, categoría A en Minciencias, en la Universidad Católica Luis Amigó, Medellín, Colombia. Correo electrónico: sebastian.diazbo@amigo.edu.co; ORCID: <https://orcid.org/0000-0001-7314-6412>.
- 3 Magíster en Derecho de la Universidad Católica Luis Amigó. Docente de Derecho Informático y Nuevas Tecnologías, Derecho Procesal Administrativo, coordinador del área de Derecho Público del Programa de Derecho y líder del grupo de investigación *Jurídicas y Sociales*, categoría A en Minciencias, en la Universidad Católica Luis Amigó, Medellín, Colombia. Correo electrónico: dany.gomezag@amigo.edu.co y danygomezagudelo@hotmail.com; ORCID: <https://orcid.org/0000-0003-2687-0146>.

Palabras clave: Acción de tutela. ChatGPT. Inteligencia artificial.

Sumario: 1. Introducción. 2. Metodología. 3. Definición y concepto de Inteligencia Artificial. 4. Concepto de jurisdicción en el derecho internacional y en Colombia. 5. Antecedentes del uso de ChatGPT en la acción de tutela. 5. Uso de la Inteligencia Artificial en la acción de tutela. 6. Conclusiones. Referencias.

1. Introducción

En Colombia, durante el 2023 se conoció por la prensa de una acción de tutela que fue resuelta por parte del Juzgado Primero Laboral del Circuito de Cartagena, la cual utilizó respuestas extraídas a través de ChatGPT para resolver la controversia judicial, con el argumento de que la ley 2213 de 2022 que implementa la justicia digital lo permite.

Esta obra despliega la pregunta de investigación: ¿cuáles son las consecuencias jurídicas de utilizar ChatGPT para resolver una la acción de tutela? Este análisis nos lleva a determinar las implicaciones de esta inteligencia artificial y conocer cuáles son sus limitaciones y dimensiones, cuando se busca proferir decisiones judiciales de forma mucho más rápida.

Este texto establece nociones jurídicas sobre el uso de ChatGPT en la justicia de Colombia, a fin de realizar un análisis de la realidad actual de la acción de tutela, más conocida como la acción de amparo en el derecho comparado. Al respecto, Gamboa (2019) señala: “la robótica, el big data y la inteligencia artificial son nuevas tendencias tecnológicas que revisten una importancia mayúscula dentro del contexto tecnológico actual” (p. 194).

Así las cosas, el objetivo general de la investigación es determinar las consecuencias jurídicas utilizar ChatGPT en la acción de tutela.

Finalmente, el corpus de esta investigación tiene dos capítulos temáticos y un acápite final de conclusiones. En el primer capítulo, se desarrolla la definición y el concepto de In-

teligencia Artificial. En el segundo capítulo, se abordó el concepto de jurisdicción en el derecho internacional y en Colombia. En el tercero, se da cuenta de los antecedentes del uso de ChatGPT en la acción constitucional. En el cuarto capítulo, se abordaron los aspectos positivos y negativos de uso al resolver una acción de tutela.

2. Metodología

En esta obra se seleccionó el enfoque cualitativo de investigación, toda vez que se recolectaron datos empíricos y fuentes documentales de alta calidad, con el fin de interpretar los datos en relación con los objetivos trazados en la investigación.

En este estudio se seleccionó la técnica de la revisión documental con miras a analizar información actual sobre la temática, a partir de revistas indexadas y libros resultado de investigación. Así mismo, se obtuvieron análisis de información extraída de bases de datos como vlex, google académico, science direct y scopus.

El paradigma escogido fue el de la investigación socio-jurídica, con un método hermenéutico. Al respecto, Duque et al (2018) afirman: “con este enfoque teórico, el investigador puede obtener el estudio de las problemáticas que detecta de la dogmática en su relación con la realidad social” (p. 23). De modo que, fueron obtenidos datos de las fuentes formales del derecho, y se realizaron análisis de las implicaciones que tuvo en la realidad la acción de tutela resuelta por el Juzgado Primero Laboral del Circuito de Cartagena que utilizó ChatGPT.

3. Definición y concepto de Inteligencia Artificial

Previo a estudiar el uso de una inteligencia artificial como ChatGPT para la toma de decisiones judiciales por parte de una inteligencia artificial, es importante definir la naturaleza misma de la herramienta a utilizar. Esto porque no se podrá determinar si es adecuado usar este programa o no para una tarea como administrar justicia, si no se conoce la forma de la tecnología a utilizar.

La inteligencia artificial ha sido desarrollada por la doctrina internacional en el ámbito de la ingeniería, pero también en el ámbito del derecho. La primera definición de la que se tenga registro fue propuesta en 1956 por J. McCarthy. Él y sus coautores propusieron que fuera la ciencia para crear máquinas o programas de computadores que pudieran emular la inteligencia humana a través de la realización de tareas que requieren habilidades especiales tales como el aprendizaje, el razonamiento y la adaptación (McCarthy, 1956).

Esta primera definición otorgada por la ciencia emulaba mucho el concepto de inteligencia artificial con el ámbito humano. El problema principal que tenía era que creía que, para desarrollar una máquina de este estilo, era necesario que realizara el mismo razonamiento que realizan los seres humanos, sin tener en cuenta que la inteligencia artificial se desarrollaría en el futuro con capacidades muy similares a la de las personas, pero, funcionando con razonamiento que difiere del proceso cognoscitivo biológico.

Pasaron muchos años sin que se llegara a un avance en el concepto de lo que es una IA. En 1987, Chandrasekaran afirmaba que existía un gran desorden en la forma como la definición de IA había sido trabajada por la humanidad. Resumía los aportes que se habían tenido hasta ese momento al afirmar que una inteligencia artificial era un sistema con la capacidad de desarrollar actividades que requirieran el uso de intelecto similar al que poseen los seres humanos (Chandrasekaran, 1987).

Otra definición que puede dar luces sobre el avance que la definición de esta tecnología iba teniendo en el tiempo, fue la que afirmaba que era la capacidad de un sistema para poder actuar de acuerdo con una circunstancia determinada. Estableciendo que este “actuar de acuerdo” era la probabilidad de éxito que tenían las acciones que tomaba el sistema. Afirmando además que la probabilidad de éxito se media a partir de los estándares comportamentales que se tenía para el sistema en situaciones determinadas (Albus, 1991).

Ya en el año 2007, los cambios en la programación tangible de máquinas con habilidades de pensamiento más complejas podían manifestarse en la forma como la definición de

Inteligencia Artificial iba cambiando. Importantes científicos afirmaron que la IA podía definirse como una cualidad de un programa para relacionarse con un ambiente determinado. Esta cualidad se refiere a la capacidad para sacar un verdadero provecho de este ambiente y esa capacidad de sacar provecho dependerá de su capacidad de adaptación a las diferentes circunstancias que ese ambiente le plantee.

A pesar de los avances que se iba teniendo en temas de inteligencia artificial y la notoria diferencia con su comportamiento humano, autores en el siglo XXI seguían entendiendo estos sistemas como la capacidad que tienen las máquinas para realizar tareas que piden una alta capacidad cognoscitiva que le es propia a los seres humanos de forma exclusiva. (Kurzweil, 2013) Este tipo de definiciones que confunden la inteligencia artificial con la humana, deben ser replanteadas al no entender que para llegar a un mismo objetivo no es necesario seguir el mismo camino. Una máquina podrá llegar un momento a realizar las mismas tareas de los seres humanos y hasta más con el uso de razonamientos abstractos, sin necesidad de utilizar la misma vía para el razonamiento abstracto que tendría un ser humano.

De acuerdo con (Barnett and Treleaven, 2018) se puede clasificar la inteligencia artificial de dos formas: la primera clasificación es aquella en la cual los códigos forman programas basados en conocimientos. Es decir, las decisiones tomadas por el programa están basadas en normas explícitas que se pueden rastrear. Se puede conocer en la totalidad las decisiones que tomó el código con base en la forma como está programado. Esta primera clasificación se subdivide a su vez en sistemas basados en reglas y en códigos base de razonamientos. En los sistemas basados en reglas, las decisiones son tomadas por un código de programación configurado en el “si, entonces luego” (*If then else*). En los sistemas base de razonamientos las decisiones van variando a medida que el programa va trabajando, es decir, se toman decisiones con el apoyo de lo aprendido en procesos primigenios.

La segunda clasificación propuesta por los autores es el aprendizaje de máquina o *machine learning*. En esta clasificación

se da un código base al sistema para que pueda funcionar, pero en lugar de darle comandos predeterminados para tomar las decisiones, el sistema va modificando su código a medida que va rastreando la información que encuentra o le es suministrada. Dentro de esta clasificación se encuentran varias formas de aprendizaje de la máquina con base en la forma en que se le brinda la información. En el aprendizaje dirigido, en el cual la los datos de los cuales el programa se alimenta son filtrados y suministrados por sus creadores; y el aprendizaje no dirigido, en el cual la máquina obtiene conocimiento aleatorio según su criterio. Esta última clasificación es la que comúnmente podría llamarse como cajas negras. Esto porque no es posible determinar los razonamientos que llevaron a la máquina a pensar de una forma o otra (Barnett and Treleaven, 2018).

Más recientemente, la inteligencia artificial ha sido definida como un sistema programado por los seres humanos, a través de tecnologías de la información y las comunicaciones, con la capacidad para: hacer recomendaciones, predicciones y tomar decisiones en ambientes reales o virtuales, de acuerdo con una serie de comandos dados por su creador (Atabek, 2023).

Luego de este breve, pero importante recorrido bibliográfico donde se exploraron las diferentes posturas sobre lo que se puede entender como inteligencia artificial, es menester proponer una conceptualización propia para este escrito, con base en lo ya dicho por los otros científicos. La inteligencia artificial puede ser definida como la capacidad que tiene una máquina para realizar una serie de tareas complejas que requieren de un razonamiento previo y abstracto para su concreción con éxito. Este razonamiento debe provenir de una serie de parámetros que la máquina obtiene luego de interactuar con distintas fuentes de información. Estas fuentes de información pueden ser suministradas de forma manual o automática a dicho sistema y son las que le permitirán tomar las decisiones acertadas para realizar con éxito las tareas a realizar.

ChatGPT, como se observará más adelante, es una herramienta de inteligencia artificial que fue programada para

poder obtener información de diversas fuentes y que le permitirá asumir la tarea de responder preguntas determinadas de acuerdo con lo solicitado por el usuario, en un lenguaje que busca emular el lenguaje escrito que utilizamos los seres humanos.

4. Concepto de jurisdicción en el derecho internacional y en Colombia

Una vez definido el concepto de inteligencia artificial, y con el propósito de mirar si un sistema inteligente puede realizar tareas de administración de justicia, es muy importante que se aborde el concepto de jurisdicción en un Estado democrático, social y de derecho como el colombiano. Solo entendiendo muy bien el concepto de IA y de jurisdicción es que se puede mirar si ambos conceptos pueden ser eventualmente compatibles.

La definición más básica de jurisdicción en el derecho internacional proviene de su máximo órgano, las Naciones Unidas. La ONU establece que la jurisdicción es la capacidad que tienen los tribunales de un país para juzgar y resolver los conflictos que se originan por un conflicto de intereses entre dos partes. Esta capacidad para dar solución a un conflicto implica la posibilidad de imponer su decisión por la fuerza legítima del estado. Esta legitimidad proviene de la capacidad que tiene un estado para dictar normas y leyes de acuerdo con los acuerdos comunes. Es por ello que es fundamental que los tribunales la hora de tomar sus decisiones se basen en el derecho (ONU, 2010).

Más adentrados en el continente americano, la Organización de Estados Americanos OEA, ha definido la jurisdicción como el poder que tienen los tribunales y sus jueces de menor categoría para conocer de los conflictos que se encuentran dentro de su competencia, a su vez, juzgar los mismos de acuerdo a la ley y siempre en garantía de los derechos humanos y fundamentales de las personas (OEA, 2002).

Ya en Colombia, la jurisdicción ha sido definida por la Constitución Nacional como “una función pública que implica

la potestad de administrar justicia en nombre del Estado. Se ejerce por los tribunales, juzgados y demás órganos que la ley determine, en la forma que esta misma establezca” (Colombia, 1991).

Importantes doctrinantes han definido la jurisdicción en Colombia como la función estatal de resolver conflictos de implicaciones jurídicas que surgen entre dos personas, aplicando las distintas fuentes de derecho de acuerdo con la Constitución y las normas internacionales de derechos humanos, buscando la efectiva protección de los derechos humanos y la justicia (De Greiff, 2010).

De acuerdo con estas definiciones, la jurisdicción es la facultad que tiene un juez para determinar la solución a un conflicto determinado que se pone en su conocimiento por la facultad que le ha sido otorgada por la ley, y que tendrá que fallar de acuerdo a la misma. Es muy importante resaltar que, si bien estas dos definiciones nos acercan mucho a la definición de lo que es jurisdicción, se vuelve menester resaltar que la jurisdicción es una facultad que tiene el Estado, pero que proviene de la legitimidad que él mismo posee. Sin legitimidad no puede haber una efectiva jurisdicción.

El que una decisión judicial pueda tener legitimidad implica el reconocimiento social de la justicia, es decir, la aceptación de la validez de las decisiones judiciales, que implica la asimilación de las mismas por sus procedimientos y el fondo mismo de las decisiones a la hora de resolver casos concretos (Bonilla, 2009).

La legitimidad de una corte proviene de la capacidad y de la autoridad moral que tiene un sistema judicial para poder tomar decisiones vinculantes en materia legal y que las mismas sean efectivamente acatadas por una sociedad (Balla, 2018). En el mismo sentido, autores como González y Galindo han afirmado que la jurisdicción es la percepción que tienen los ciudadanos acerca de la autoridad de un tribunal para resolver un conflicto aplicando la legislación vigente, esta autoridad se fundamenta en sus principios de imparcialidad, independencia y competencia por parte de los jueces (González and Galindo, 2019).

De acuerdo con la Corte Constitucional de Colombia, la legitimidad judicial proviene de la capacidad de los órganos judiciales para inspirar la confianza y el respeto, mediante la observancia de las reglas procesales y sustanciales y en la debida aplicación del ordenamiento jurídico vigente (Sentencia T-568/08).

En sentencias similares, la Corte también ha afirmado que la legitimidad de la justicia está dada por la confianza que tienen los ciudadanos en las instituciones encargadas de administrar justicia en Colombia. Esta confianza se funda en la vigencia de un conjunto de valores tales como la imparcialidad, la independencia, la transparencia, la idoneidad y la eficiencia en la toma de decisiones (Sentencia T-724, 2003).

Luego de observar todas estas definiciones sobre la jurisdicción y la legitimidad que tiene que tener la misma para pueda operar de manera óptima y pueda efectivamente materializar sus decisiones en el mundo real, es muy importante plantear la pregunta ¿puede un sistema de inteligencia artificial gozar de la misma legitimidad de la que goza un juez de carne y hueso que ha vivido en un país y sociedad determinada?

Es muy importante recordar que la legitimidad que tienen los jueces no proviene únicamente de la ley. Como se pudo observar en las definiciones, la legitimidad tiene un alto índice de autoridad moral por parte de las autoridades que están aplicando una ley determinada.

Cuando un ciudadano se acerca a la justicia para pedir la garantía de sus derechos constitucionales y legales ¿Qué tan satisfecho quedará con la decisión de un sistema de inteligencia artificial que ni siquiera sabe muy bien como tomó su decisión? En el caso de un juez real, él puede observar que es una persona como él, que vive en su misma ciudad y que la única diferencia que puede tener con él es que el juez ha decidido dedicar su vida al estudio de la ley y del derecho para poder estar donde está sentado y darle una solución a su problema. Es muy improbable que un sistema de inteligencia artificial como ChatGPT pueda inspirar el mismo respeto, y por lo tanto, las decisiones del sistema judicial podrían estar en grave peligro de tener un desacatamiento por parte de los administrados.

No se puede cerrar esta obra sin antes señalar la importancia que puede tener la experiencia propia, como ciudadano y persona, de un juez en una sociedad determinada, frente a la que puede tener un sistema de inteligencia artificial como ChatGPT en la lectura de documentos en internet.

Un juez ha vivido toda su vida en la sociedad sobre la cual tomará las decisiones que está tomando. Ha sido una persona que nació en dicho país (en la gran mayoría de los casos se pide que la persona sea un nacional de allí) o por lo menos ha vivido el tiempo suficiente para conocer cuáles son los valores que rigen a una sociedad determinada. La aplicación del derecho no es la simple aplicación mecánica de la ley, la escuela exegetica del derecho hace ya varios siglos que fue superada. La aplicación del derecho es la aplicación de la justicia, y para llegar a una verdadera justicia se hace imperativo que un operador judicial tome decisiones basado también en los principios morales y éticos que rigen una sociedad (lo que muchos doctrinantes han decidido llamar *ius naturalismo*).

Un sistema de inteligencia artificial programado por una empresa localizada en otro país tal vez tenga la capacidad de procesar gran cantidad de documentos en muy poco tiempo, pero jamás tendrá lo que un operador de judicial real, un gran tiempo de interacción en una sociedad que le ha permitido conocer la verdadera manifestación de lo que es la justicia.

5. Antecedentes del uso de ChatGPT en la acción de tutela

En 2023, un juzgado ubicado en Cartagena, Colombia, falló una acción de tutela acudiendo a ChatGPT. El caso se trató de una solicitud de exoneración del pago de la cuota moderadora a la EPS, de un niño que padecía del trastorno de autismo.

Con el fin de resolver la acción constitucional, el juez expuso varios interrogantes con problemas jurídicos del caso y copió y pegó la respuesta en el acápite de consideraciones o motivación de la decisión judicial. A continuación, se presenta una imagen del fallo de tutela:

Figura 1: Sentencia del Juzgado Primero Laboral del Circuito de Cartagena.

Por otra parte, atendiendo que la Ley 2213 de 2022 tiene por objeto la incorporación de las TIC en los procesos judiciales., se resolverá extender los argumentos de la decisión adoptada, conforme al uso de inteligencia artificial (IA). Por consiguiente, se procedió a incluir apartes del problema jurídico planteado en la presente sentencia y se ingresó en el aplicativo <https://chat.openai.com/chat>.

Se hicieron en el aplicativo las siguientes preguntas.

¿Menor autista esta exonerado de pagar cuotas moderadoras en sus terapias?

La respuesta del aplicativo IA:

Sí, es correcto. De acuerdo con la normativa en Colombia, los menores con diagnóstico de autismo están exonerados de pagar cuotas moderadoras en sus terapias. Este beneficio está regulado por la Ley 1753 de 2015 y se aplica a todas las instituciones prestadoras de servicios de salud públicas o privadas que reciban recursos del sistema general de seguridad social en salud.

¿Las acciones de tutela en estos casos se deben conceder?

La respuesta del aplicativo IA:

“En caso de que una institución prestadora de servicios de salud no cumpla con la exoneración de cuotas moderadoras para un menor con diagnóstico de autismo, la persona o sus representantes legales pueden presentar una acción de tutela para proteger sus derechos. La tutela es un mecanismo judicial rápido y eficaz para proteger derechos fundamentales que se ven vulnerados.

Si se demuestra que el derecho a la salud del menor con autismo está siendo afectado por la exigencia de pagar cuotas moderadoras, es probable que la acción de tutela sea concedida y se ordene a la institución prestadora de servicios de salud cumplir con la exoneración. Sin embargo, cada caso es único y la decisión final depende de los hechos y circunstancias específicas del caso.”.

Fuente: Sentencia del Juzgado Primero Laboral del Circuito de Cartagena, radicado 13001410500420220045901, 30/01/2023.

El fallo del juez hizo énfasis en que la Ley 2213 de 2022 no prohibió el uso de la inteligencia artificial en el país, y que su finalidad al acudir a ChatGPT era reducir el tiempo de los empleados públicos en la escritura de fallos de tutela, así mismo, que corroboró la información brindada por ChatCPT.

El juez expresó que verificó la veracidad de la información emitida por la inteligencia artificial. En todo caso, en la opinión pública se generaron dudas de si el Congreso de Colombia realmente permitió el uso de la inteligencia artificial, también cuáles son sus efectos legales, los aspectos positivos y negativos. Al respecto, Juca-Maldonado (2023) afirma:

No se pueden detener los adelantos tecnológicos y el cambio que estos producen en la vida de las personas y las distintas áreas de las ciencias o actividades que afectan o modifican, solo queda evolucionar, adaptarse a los cambios y replanearse la forma como se educa y aprende, que cada vez está al alcance de más personas (Juca-Maldonado, 2023, p. 8).

Además, en relación a lo señalado en la acción de tutela que menciona la posibilidad de utilizar la inteligencia artificial, en la jurisdicción ordinaria, el Código General del Proceso estableció en su artículo 103 señala que: “en todas las actuaciones judiciales deberá procurarse el uso de las tecnologías de la información y las comunicaciones en la gestión y trámite de los procesos judiciales, con el fin de facilitar y agilizar el acceso a la justicia” (Ley 1564, 2012, art. 103).

En este orden de ideas, el Congreso de Colombia expidió la ley 2213 de 2022, que regula los procesos virtuales en la actualidad. Luego de efectuar un análisis de esta ley, se detecta que en ningún artículo se permitió expresamente el uso de la IA, no obstante, tampoco se prohibió. Por ello, a partir de una interpretación de la finalidad que tuvo el legislador con esta ley, es posible señalar que si se permite a los jueces de Colombia utilizar herramientas de IA como ChatGPT, sobre el particular el legislador señaló:

Se podrán utilizar las tecnologías de la información y de las comunicaciones, cuando se disponga de los mismos de manera idónea, en la gestión y trámite de los procesos judiciales y asuntos en curso, con el fin de facilitar y agilizar el acceso a la justicia (Ley 2213, 2022, art. 2).

En este orden de ideas, Generative Pre-training Transformer más conocido como ChatGTP, fue desarrollada por la empresa OPEN AI, con el propósito de brindar de cualquier

tema, con características de respuesta muy parecidas a las humanas y bastante amable a la hora de responder, dado que “el sistema está basado en un modelo de lenguaje generativo creado con 175 millones de parámetros y entrenado a partir de 8 millones de documentos” (Hughes, 2023).

De modo que, ChatGPT consiste en una herramienta creada por OpenAI, una organización que desarrolla inteligencia artificial. Su uso para todas las personas que tengan acceso a internet “ha generado gran entusiasmo y controversia dado que es uno de los primeros modelos que puede conversar de manera convincente con sus usuarios en diferentes idiomas sobre una amplia gama de temas” (Ulloa, 2023, p. 1).

ChatGPT se ocupa de brindar respuestas a partir de la información que aprende y procesa durante su periodo de entrenamiento. Sobre el particular, Corvalán (2018) afirma:

Hay que trabajar sobre la siguiente cuestión: cuánta intervención del ser humano resulta necesaria para que el resultado del procesamiento de información y de los datos de sistemas de IA sea legítimo, respetuoso y promotor de la efectividad de los derechos de las personas (Corvalán, 2018, p. 20).

De esta manera, ChatGPT puede continuar aprendiendo con el tiempo, es decir, puede detectar sus propios errores, tiene la facultad de emitir respuestas más útiles y acertadas. Esencialmente, esta es la labor de sus desarrolladores: mejorar su habilidad para adaptarse a los tipos de preguntas y proferir respuestas coherentes, ante una extensa variedad de posibles elecciones de respuestas.

6. Uso de la Inteligencia Artificial en la acción de tutela

Esta herramienta tiene la capacidad para emitir todo tipo de respuestas, aspecto que llama la atención de los internautas, es necesario tener ciertas precauciones al utilizar ChatGPT, sobre todo, cuando los jueces deban resolver una acción de tutela.

ChatGPT no tiene sentimientos. No diferencia el estado de ánimo o la personalidad del usuario que le escribe. Tampoco puede diferenciar entre la veracidad o inexactitud de los datos que emite. Simplemente profiere información a partir de una diversidad de fuentes en varios idiomas, que provienen de páginas web, artículos, libros, información periodística. Sobre el particular, Díaz y Bustamante (2022) explican: “la inteligencia artificial es la capacidad que tiene un software para tomar decisiones y aprender sin un programa explícito de comandos en su código de programación” (p. 18).

Si bien es una inteligencia artificial avanzada, aún hay áreas en las que su comprensión puede ser limitada, y siempre hay posibilidades para mejorar su funcionamiento, por ejemplo, en conocimientos de ingeniería, medicina o en el derecho, sus respuestas no son 100% confiables, por lo que muchas veces sugiere acudir a un experto tenga mayor conocimiento que ChatGPT.

En relación a su uso para proferir sentencias, aunque esta inteligencia artificial puede brindar información general sobre las normas que rigen un proceso, la generación de decisiones judiciales requiere de una mayor comprensión mucho más profunda que conocer el marco legal, con la capacidad de tener en cuenta diversos criterios y circunstancias del caso que incidan en la decisión, este es uno de los aspectos en el que ChatGPT no tiene este alcance. Sobre el particular, Morales (2021) afirma: “el progreso de la IA no puede ir en desmedro de los derechos fundamentales de las personas, tanto naturales como jurídicas” (p. 33).

De esta manera, el juez, al momento de proferir sentencia, está en la obligación de conocer la normativa que rige el proceso, y no depender exclusivamente de la tecnología. Así mismo, debe acudir a las fuentes formales del derecho y literatura especializada, para adoptar decisiones con fundamentos jurídicos y de forma coherente con la legislación colombiana.

La investigación detectó otra debilidad de ChatGPT relacionada con que su conocimiento es limitado hasta septiembre de 2021, por ejemplo, si se le pregunta sobre quién es el presidente de Colombia, su respuesta no es acertada. Lo

anterior, porque cualquier evento que haya ocurrido después de esa fecha, no estará incluido en su entrenamiento. De esta manera, si dado que ChatGPT no cuenta con las modificaciones efectuadas por la ley o la jurisprudencia posterior a la fecha señalado, es muy probable que brinde una respuesta incorrecta o incompleta.

Así mismo, se destaca que tiene una incapacidad para comprender la totalidad del texto, es decir, se pueden presentar dificultades para entender el contexto más amplio, matices del problema, incluyendo circunstancias concretas del accionante o accionado (demandante o demandado), lo cual podría afectar la objetividad y veracidad de la respuesta.

A pesar de las críticas expuestas en la presente investigación en relación con el uso de ChatGPT, no se pretende señalar que se deba prohibir el uso de la herramienta, dado que este modelo de lenguaje también puede traer beneficios positivos, como mejorar la creatividad y la generación de ideas novedosas. Al respecto, López (2022) afirma:

Es posible la identificación de los datos clave y revisión ágil de los procedimientos, transcripción en texto de vistas orales, realizar trabajos repetitivos de forma precisa y rápida, relacionar la legislación que afecta a expedientes similares, determinar su naturaleza jurídica y proponer fundamentos jurídicos a aplicar, apoyo a la toma de decisiones judiciales (López, 2022, p. 96).

ChatGPT puede procesar grandes cantidades de datos de forma más rápida, lo que podría ayudar a adoptar decisiones más pronto, aspecto que reclama la Constitución Política de Colombia al establecer que: “la función administrativa está al servicio de los intereses generales y se desarrolla con fundamento en los principios de igualdad, moralidad, eficacia, economía, celeridad, imparcialidad y publicidad” (Const, 1991, art. 209).

Además, la inteligencia artificial tiene la capacidad de aplicar los principios jurídicos de forma uniforme, de esta forma se garantiza que la decisión judicial cumpla con el principio de imparcialidad y transparencia.

De este modo, la materialización de estos beneficios depende en gran medida de que los jueces aprendan a utilizar adecuadamente este modelo de lenguaje y de cómo se integre al proceso para agilizar la expedición de la decisión, a través de la acción de tutela. Al respecto, Galvis et al (2022) expone:

La opción que mejor se adecua a las necesidades del sistema judicial colombiano es la aplicación de IA con un modelo asistencialista mixto en el que el juez mantiene el control del proceso. De esta forma, se reducen los sesgos, la inflexibilidad del precedente judicial y se mantiene la obligación de motivar adecuadamente las decisiones judiciales (p. 21).

Finalmente, ChatGPT no debe reemplazar al juez constitucional, sino por el contrario, se puede utilizar como una herramienta complementaria para adoptar decisiones que involucren la protección de los derechos fundamentales, de una forma más rápida, evitando sacrificar los principios del proceso jurisdiccional.

7. Conclusiones

A pesar de que ChatGPT puede proporcionar información sobre aspectos legales, no cuenta con el alcance suficiente para resolver directamente controversias sobre derechos fundamentales a través de la acción de tutela. Es necesario que tanto jueces como abogados cuenten con un conocimiento profundo y actualizado de la normativa y jurisprudencia que rige el caso, teniendo en cuenta las circunstancias fácticas, para que se evite adoptar una decisión contraria al ordenamiento jurídico. En este sentido, ChatGPT sí puede ser utilizada como una herramienta complementaria para que los jueces, al momento de resolver una acción de tutela adopten sus decisiones, pero no debe ser la única fuente de información, ya que el juez debe aplicar las fuentes formales del derecho, y tener en cuenta los cambios legales o jurisprudenciales, que no los podría brindar la versión gratuita de ChatGPT.

Es fundamental comprender que esta herramienta puede cometer muchos errores, que pueden llevar al juez a tomar

una decisión que afecte los derechos fundamentales de los ciudadanos y por lo tanto, será siempre su deber al usar este tipo de tecnologías, verificar la información que es brindada por el sistema de inteligencia artificial, dado que los beneficios dependen en de que se utilice apropiadamente la herramienta y de cómo se vincule al proceso para emitir un fallo.

Es importante que las decisiones judiciales en un futuro no sean tomadas en su totalidad por sistemas de inteligencia artificial. A la hora de tomar decisiones, no solo se está teniendo en cuenta la aplicación efectiva de un sistema judicial, si no que las decisiones deben contar con un respaldo de legitimidad que es muy difícil que un software desarrollado por una empresa multinacional pueda tener por encima de un ciudadano del país donde se está aplicando el fallo. Un sistema de inteligencia artificial tal vez pueda tener una más alta capacidad de procesamiento de información, pero no podrá tener el respaldo y la experiencia social que se necesitan para tomar decisiones que puedan solucionar óptimamente los conflictos sociales.

Es necesario que el Consejo Superior de la Judicatura, órgano director de la Rama Judicial en Colombia, brinde mayor capacitación a todos los jueces en cómo utilizar de forma eficaz y responsable la tecnología, permitiendo su uso de forma informada y ética, a fin de garantizar los derechos fundamentales de las personas, ante el uso de la inteligencia artificial.

Referencias

- Albus, J. S. (1991) 'Outline for a theory of intelligence', *IEEE Transactions on Systems, Man and Cybernetics*, 21, pp. 473-509.
- Atabekov, A. (2023) 'Artificial Intelligence in Contemporary Societies: Legal Status and Definition, Implementation in Public Sector across Various Countries', *Social Sciences*, 12(3), p. 178.

- Balla, H. (2018) 'Building Legitimacy for Domestic and International Courts: A Case for Incorporating Deliberative Democracy', *Global Constitutionalism*, 7(1), pp. 42-67.
- Barnett, J. and Treleaven, P. (2018) 'Algorithmic Dispute Resolution—The Automation of Professional Dispute Resolution Using AI and Blockchain Technologies', *Computer Journal*, 61(3), pp. 399-408.
- Bonilla, D. (2009) 'La legitimidad judicial'. *Anuario de Derecho Constitucional Latinoamericano*, 15, pp. 243-257 [online]. Available at: <https://dialnet.unirioja.es/servlet/articulo?codigo=4009024>.
- Chandrasekaran, B. (1987) 'What kind of information processing is intelligence? Perspective on AI paradigms and proposal', *Paper presented at the Second International Symposium on Methodologies for Intelligent Systems*, Charlotte, NC, USA, pp. 14-17; pp. 43-44.
- Colombia (1991) *Constitución Política de Colombia*. 4 de julio de 1991.
- Colombia (2012) *Ley 1564. Julio 12 de 2012*. Código General del Proceso [online]. Available at: http://www.secretariasenado.gov.co/senado/basedoc/ley_1564_2012.html (Accessed: 14 August 2023).
- Colombia (2013). *Ley 2213. Junio 13 de 2022*. Ley de justicia digital [online]. Available at: http://www.secretariasenado.gov.co/senado/basedoc/ley_2213_2022.html (Accessed: 14 August 2023).
- Comisión Interamericana de Derechos Humanos. (2002). *Informe sobre el acceso a la justicia en las Américas* [online]. Available at: https://www.oas.org/es/cidh/defensoria/acceso_justicia/Informe%20acceso%20a%20la%20justicia%20en%20las%20Americas.pdf (Accessed: 14 August 2023).
- Corte Constitucional de Colombia (Sala Primera de Decisión). (2008) *Sentencia T-568 de 2008, del 29 de mayo* [online]. Available at: <https://www.corteconstitucional.gov.co/relatoria/2008/T-568-08.htm> (Accessed: 14 August 2023).

- Corte Constitucional de Colombia (Sala Primera de Revisión). (2003). *Sentencia T-724 de 2003, 20 de agosto* [online]. Available at: <https://www.corteconstitucional.gov.co/relatoria/2003/t-724-03.htm> (Accessed: 14 August 2023).
- Corvalán, J. G. (2018) ‘Inteligencia artificial: retos, desafíos y oportunidades – Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia’, *Revista de Investigações Constitucionais, Curitiba*, 5(1), pp. 295-316 [online]. Available at: <https://doi.org/10.5380/rinc.v5i1.55334> (Accessed: 14 August 2023).
- De Greiff, P. (2010) *Jurisdicción y derechos humanos*. Bogotá: Universidad de los Andes.
- Díaz Bolívar, S. and Bustamante Rúa, M. M. (2023) ‘Análisis de los sistemas de solución de conflictos en línea en el mundo y propuesta para su implementación en Colombia’, *Revista Chilena De Derecho Y Tecnología*, 11(2), pp. 245-276.
- Duque, S., González, F., Cossio, N. and Martínez, S. (2018) *Investigación en el saber jurídico*. Medellín: Universidad de Antioquia.
- Galvis Vega, G., Alfonso Acosta, G., Rodríguez Peñaloza, A., Cabeza Zambrano, C., Olarte Mojica, L., Garzón Fierro, V., Chaves Rodríguez, P. and Trespacios Rojas, I. (2022) *Inteligencia artificial en el proceso de predicción de decisiones judiciales como modelo asistencialista mixto: entre la codificación y la justicia equitativa* [online]. Available at: <https://repositorio.uniandes.edu.co/handle/1992/59884> (Accessed: 14 August 2023).
- Gamboa, R. (2019) ‘Algunas perspectivas, desafíos y aplicación de la tecnología en la profesión legal’, *Tecnologías al servicio de la Justicia y el Derecho*, Pontificia Universidad Javeriana, pp. 179-198.
- González-Pérez, J. G. and Galindo-Aldana, J. A. (2019) ‘Legitimidad jurisdiccional y confianza en las instituciones judiciales Una revisión crítica de la literatura’, *Revista de Derecho de la Universidad Francisco de Vitoria*, 31, pp. 145-165.

- Hughes, A. (2023) 'ChatGPT: Everything you need to know about OpenAI's GPT tool BBC', *Science focus* [online]. Available at: <https://www.sciencefocus.com/future-technology/gpt-3/> (Accessed: 14 August 2023).
- Juca-Maldonado, F. (2023) 'El impacto de la inteligencia artificial en los trabajos académicos y de investigación' *Revista Metropolitana de Ciencias Aplicadas*, 6(S1), pp. 289-296 [online]. Available at: <http://remca.umet.edu.ec/index.php/REMCA/article/view/656/662> (Accessed: 14 August 2023).
- Kurzweil, R. (2013) *How to Create a Mind: The Secret of Human Thought Revealed*. New York: Penguin.
- López, Y. (2022) 'Transformación digital de la justicia y discapacidad' in Bustamante, M., Henao, A. and Toro, L. (edn.), *Justicia y Sociedad 5.0*. Envigado, Colombia: Institución Universitaria de Envigado, pp. 89-109.
- McCarthy, J., Minsky, M., Rochester, N. and Shannon, C. E. (1956) *Proposal for the Dartmouth Summer Research Project on Artificial Intelligence* [online]. Available at: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html> (Accessed: 14 August 2023).
- Morales Cáceres, A. (2021) 'El impacto de la inteligencia artificial en el Derecho', *Advocatus*, 39, pp. 39-71.
- Organización de las Naciones Unidas. (2010) *Directrices básicas sobre el derecho a un recurso efectivo y a reparación en casos de violaciones manifiestas de las normas internacionales de derechos humanos y de violaciones graves del derecho internacional humanitario* [online]. Available at: <https://www.ohchr.org/SP/ProfessionalInterest/Pages/RemedyAndReparation.aspx> (Accessed: 14 August 2023).
- Organização dos Estados Americanos - OEA (2022) *Informe anual de la Comisión Interamericana de Derechos Humanos*, Available at: <https://www.cidh.oas.org/annualrep/2002sp/indice.htm> (Accessed: 12 September 2023).
- Ulloa, G. (2023) 'El desafío del uso de inteligencia artificial para la elaboración de la literatura científica: el caso de ChatGPT, un debate abierto', *Cuadernos Médico Sociales*, 63(1), pp. 27-31.

SEGURANÇA PÚBLICA EM CIDADES INTELIGENTES SOB UMA PERSPECTIVA LOCAL

José Jorge Barreto Torres¹

Patrícia Verônica Nunes Carvalho Sobral de Souza²

Diogo de Calasans Melo Andrade³

DOI: 10.29327/5312711.1-10

Resumo: Expansão urbana acelerada e consumo destrutivo constituem uma fórmula desastrosa para os momentos presente e futuro da humanidade. As Cidades Inteligentes se mostram como um farol de esperança na construção utópica de uma sociedade tecnológica e sustentável. O aproveitamento das Tecnologias de Informação e Comunicação (TICs) nas tarefas de gestão municipal, nas dimensões de segurança, educação, transportes e infraestrutura pública, além da observância de técnicas e metodologias sustentáveis, podem tornar essa manifestação de sociedade evoluída em algo possível de alcançar. No eixo da Segurança Pública, diversas ideias que atuam de maneira direta e indireta nesse escopo são implantadas ao redor do planeta, com a colaboração das TICs associadas aos benefícios resultantes da sustentabilidade, mitigando os efeitos do crescimento social predatório. A utilização de drones, o compartilhamento de câmeras privadas de monitoramento,

- 1 Doutorando em Direitos Humanos pela Universidade Tiradentes. Mestre em Ciência da Computação pela Universidade Federal de Sergipe e Especialista em Bancos de Dados pela Universidade Tiradentes. Bacharel em Direito e Sistemas de Informação pela Universidade Tiradentes. E-mail: doutorado_jorgebt@souunit.com.br.
- 2 Pós-Doutora em Direito pela Mediterranea International Centre for Human Rights Research della Università Mediterranea di Reggio Calabria (Itália). Pós-Doutora em Direito e Doutora em Direito Público pela Universidade Federal da Bahia-UFBA. Professora Titular de Graduação e do Programa de Pós-graduação da Universidade Tiradentes. E-mail: patncss@gmail.com.
- 3 Doutor em direito político e econômico pela Universidade Mackenzie. Mestre em direito, na área de constitucionalização do direito pela UFS. Professor titular da graduação e do mestrado e doutorado em direitos humanos do PPGD-UNIT. Líder do grupo de pesquisa “Novas tecnologias e o impacto nos Direitos Humanos” do mestrado em direito Humanos da UNIT, advogado. E-mail: contato@profdiogocalasans.org.

os investimentos em iluminação pública, a utilização de centrais operacionais integrando os mecanismos de segurança e vigilância, tecnologias de reconhecimento facial e detecção autônoma de veículos roubados e furtados são algumas soluções empregadas para otimização dos serviços de segurança em Cidades Inteligentes. Diante desta problemática, busca-se o entendimento dos principais conceitos envolvidos no âmbito das Cidades Inteligentes para mapear sua efetivação em um cenário local. Em direção a este objetivo, foi performada uma pesquisa básica, qualitativa, exploratória, bibliográfica e documental, em busca de conceitos que servem de base para o entendimento do tema e de casos reais, extraído de cidades de referência na América Latina, para finalmente explorar o que existe em termo de políticas públicas na cidade de Aracaju, no âmbito das Cidades Digitais, aplicados especificamente à Segurança Pública. Finalmente, nota-se uma pouca aplicação dos investimentos oriundos da União, na utilização de projetos de digitalização do município sendo observada uma pequena evolução nesse tema ao longo das décadas.

Palavras-chave: Cidades inteligentes. Segurança pública. Sustentabilidade.

Sumário: 1. Introdução. 2. Cidades inteligentes: digitais e sustentáveis. 3. Políticas de segurança pública em cidades inteligentes. 4 Análise do cenário local. 5. Considerações finais. Referências.

1. Introdução

Em face do rápido crescimento populacional nos grandes centros urbanos, com o consumo dos recursos naturais de modo desenfreado e por muitas vezes predatório, surge um cenário caótico preocupante para as gerações atuais e futuras. A própria organização social tradicional não consegue acompanhar a velocidade das mudanças. Para isso, conta-se com o emprego das Tecnologias da Informação e Comunica-

ção (TICs) para automatizar diversas tarefas da gestão pública. Além disso, a modernização dos meios de comunicação, desde que feita de maneira inclusiva, encurta caminhos entre as pessoas e os serviços municipais.

Para uma cidade se tornar digital, é necessário adotar tecnologias de infraestrutura de base para apoiar os diversos mecanismos de inovação. A utilização de conectividade sem fio de maneira gratuita, o fornecimento de serviços em portais e aplicativos para dispositivos móveis são alguns dos recursos aderentes à transição de uma cidade de um meio analógico para o meio digital.

De fato, mostra-se imprescindível a utilização das TICs para a modernização desses serviços. Entretanto, deve-se levar em consideração também durante esses processos as preocupações com os dias futuros, onde atuam outras variáveis que influenciam direta ou indiretamente no bem-estar social, como a utilização equilibrada e responsável dos recursos naturais, por meio de adoção de sistemas renováveis, descarte correto de lixo, saneamento básico e qualquer outro fator que possa influenciar no desenvolvimento humano saudável.

Sob esse aspecto, os Objetivos de Desenvolvimento Sustentável (ODS) se aliam às metas do programa brasileiro voltado à sustentabilidade dos centros urbanos – o Programa Cidades Sustentáveis – que possui doze eixos temáticos subdividindo as diversas áreas de gestão municipal, detalhando os direcionamentos para que cada nicho alcance os resultados pretendidos.

Nessa toada, projetos de Cidades Inteligentes urgem como um apontamento para mitigação dos efeitos adversos provenientes do crescimento urbano desordenado e do consumo desenfreado e irresponsável dos recursos naturais. Em grande parte da literatura especializada, as Cidades Inteligentes apresentam-se como uma convergência conceitual das Cidades Digitais com as Cidades Sustentáveis, fornecendo o melhor para o bem-estar da sociedade com a utilização de inovações tecnológicas, de maneira a facilitar o dia a dia da população aliado a um consumo consciente dos elementos naturais e artificiais.

No eixo da Segurança Pública, são muitos os projetos direcionados ao melhor uso da tecnologia no combate à criminalidade, que iniciam por uma boa iluminação pública e outros fatores relacionados diretamente com esse tipo de serviço fornecido pelo Estado, como o compartilhamento de câmeras privadas com o poder público, utilização de drones para monitoramento, centrais operacionais e de inteligência para segurança, reconhecimento facial no combate ao crime, detecção automática de sinais identificadores para localização de veículos automotores roubados ou furtados, entre outras aplicações das TICs em prol do aumento da sensação de segurança.

O presente estudo objetiva apresentar o que há de concreto em projetos de infraestrutura e iniciativas voltadas às Cidades Inteligentes no município de Aracaju-SE. Com o olhar voltado ao cumprimento dessa tarefa, serão conceituados os tipos de cidades apresentadas como foco dessa temática e que se complementam: as Cidades Sustentáveis, Digitais e Inteligentes. Em seguida, enfatiza-se a busca por mecanismos inteligentes aplicados às políticas de Segurança Pública em municípios de destaque na América Latina para, finalmente, relatar o que o município de Aracaju já colocou em prática a respeito da aplicabilidade dessas tecnologias.

Para o melhor aproveitamento desse trabalho, será utilizado o método indutivo, confrontando todo arcabouço conceitual estudado para chegar ao resultado do panorama local, executando-se uma pesquisa de natureza: básica, gerando conhecimento a partir de suas descobertas; qualitativa e exploratória, ao buscar exemplos no meio acadêmico de aplicação de políticas inteligentes voltadas à segurança pública; bibliográfica, por estar amparada em livros, artigos e revistas científicas; e documental, pelo acesso às informações presentes em notícias e reportagens disponibilizadas por veículos de comunicação consolidados.

Segue um breve roteiro do que é explicitado neste artigo: No segundo capítulo são priorizados os conceitos de Cidades Sustentáveis, Digitais e Inteligentes, bem como a correlação entre seus significados. No capítulo seguinte há uma

amostra de recursos utilizados em Cidades Inteligentes voltados às políticas de Segurança Pública em cidades da América Latina são apresentados para melhor conhecimento do que pode ser explorado nesse nicho. O quarto capítulo visa trazer ao leitor, sem esgotar o assunto, o que se dispõe em matéria de infraestrutura e implantação, tanto no âmbito das Cidades Digitais quanto das Cidades Inteligentes na capital sergipana, com foco na Segurança Pública. No último capítulo são feitas as conclusões, descobertas e contribuições desta pesquisa.

2. Cidades inteligentes: digitais e sustentáveis

Antes de explorar o cerne deste estudo é necessário apurar os principais conceitos que amparam a presente linha temática. A partir da ideia a ser desenvolvida, de que a Cidade Inteligente resulta de uma relação simbiótica entre Cidade Digital e Sustentável, é mandatório aprofundar-se nestes conceitos para um melhor entendimento.

2.1. Cidade sustentável

A adoção de medidas sustentáveis, de maneira geral, objetiva mitigar os efeitos adversos sobre a utilização dos recursos naturais. Essa utilização, aliada a uma crescente populacional no planeta, representa uma fórmula destrutiva. A partir desse cenário, é preciso fazer um acompanhamento de índices variados sobre ações que causam impacto direto e indireto sobre todo ecossistema para atuar de maneira proativa, com pensamento nas gerações atuais e futuras. O resultado de todo esse esforço se traduz em um planejamento urbano saudável, aproveitando a utilização de novas capacidades tecnológicas, com o reaproveitamento e utilização equilibrada dos recursos fornecidos pela natureza (Botton et al, 2021, pp. 627-628).

Uma pesquisa realizada por Fabris et al (2020, pp. 217-218) apresenta outros conceitos ligados à Cidade Sustentável, como

[...] cidades verdes, cidades digitais, cidades inteligentes, cidades da informação, cidades do conhecimento, cidades resilientes, cidades de baixo carbono, cidades habitáveis, ecocidades, ecocidades de baixo carbono, ecocidades onipresentes, entre outros.

O grande desafio apontado por esses sistemas diversos, de acordo com os autores, é integrar o crescimento econômico a processos urbanos sustentáveis, transformando as cidades em lugares melhores de se viver. Tudo converge em políticas públicas para conscientização da população, com a participação indispensável de diferentes atores da sociedade exercendo um protagonismo de agentes transformadores.

O conceito de desenvolvimento urbano com sustentabilidade no Brasil é amparado pelo Programa Cidades Sustentáveis, que subdivide as áreas de gestão em 12 eixos incorporando as dimensões social, ambiental, econômica, política e cultural, conforme descritos na Tabela 1 (Programa Cidades Sustentáveis, 2016).

Tabela 1. Eixos da Plataforma Cidades Sustentáveis

Eixo	Objetivo Geral
Governança	Fortalecer processos de decisão pela promoção dos instrumentos da democracia participativa, fomentando a igualdade em suas múltiplas dimensões.
Bens naturais comuns	Assumir plenamente as responsabilidades para proteger, preservar e assegurar o acesso equilibrado aos bens naturais comuns.
Equidade, justiça social e cultura de paz	Promover comunidades inclusivas e solidárias.
Gestão local para sustentabilidade	Implementar uma gestão responsável, participativa e eficiente que envolva as etapas de planejamento, execução e avaliação.
Planejamento e desenho urbano	Reconhecer o papel estratégico do planejamento e do desenho urbano na abordagem sistêmica das questões ambientais, sociais, econômicas, culturais e da saúde, para benefício de todos.

Cultura para a sustentabilidade	Desenvolver políticas culturais que respeitam e valorizam a diversidade cultural, o pluralismo e a defesa do patrimônio natural, construído e imaterial. Ao mesmo tempo, promovem a preservação da memória e a transmissão das heranças naturais, culturais e artísticas, assim como, incentivem uma visão aberta de cultura, em que valores solidários simbólicos e transculturais estejam ancorados em práticas dialógicas, participativas e sustentáveis
Educação para a sustentabilidade e qualidade de vida	Integrar, na educação formal e não formal, valores e habilidades para um modo de vida sustentável e saudável.
Economia local dinâmica, criativa e sustentável	Apoiar e criar as condições para uma economia local dinâmica e criativa, que garanta o acesso ao emprego sem prejudicar o ambiente.
Consumo responsável e opções de estilo de vida	Adotar e proporcionar o uso responsável e eficiente dos recursos e incentivar um padrão de produção e consumo sustentáveis.
Melhor mobilidade e menos tráfego	Promover a mobilidade sustentável, reconhecendo a interdependência entre os transportes, a saúde, o ambiente e o direito à cidade.
Ação local para a saúde	Proteger e promover a saúde e o bem-estar dos nossos cidadãos.
Do local para o global	Assumir as responsabilidades globais pela paz, justiça, equidade, desenvolvimento, sustentável, proteção ao clima e a biodiversidade.

Fonte: Programa Cidades Sustentáveis (2016).

O referido programa atua na busca pelo equilíbrio do desenvolvimento econômico com a sustentabilidade ambiental e justiça social, analisando essa questão de forma sistêmica e horizontal, reduzindo os efeitos de mal-estar social provocados por uma crise socioeconômica mundial. O programa também se adaptou aos Objetivos de Desenvolvimento Sustentável (ODS), associando seus eixos com os 17 ODS e suas metas e

indicadores de desempenho (Programa Cidades Sustentáveis, 2016, pp. 4-5).

2.2. Cidade digital

Conectadas e interligadas com a oferta de serviços pelo meio digital, fornecidas pela implementação das mais diversas tecnologias da informação e meios de comunicação, unindo órgãos participantes na gestão pública, com a disponibilização de pontos de acesso à internet para toda população. Estas são algumas das ideias do que deve ser uma Cidade Digital e do porquê de sua existência. A ideia é a modernização dos aparelhos públicos e a inclusão digital da população que passa a acessar os serviços de maneira prática e rápida, explorando novas formas de comunicação e integração auxiliadas pelas inovações tecnológicas (Andrade and Meira, 2021, pp. 71-73).

Nesse mesmo caminho, Andrade e Cruz (2019, p. 4) apresentam as Cidades Digitais como “[...] sistemas que procuram incrementar a qualidade de vida e o conforto do ser humano por sistemas interativos com alto nível de usabilidade e flexibilidade que atendam às necessidades dos cidadãos”.

De forma semelhante, Lemos (2007, p. 142) define as Cidades Digitais ou Cibercidades como projetos que “[...] visam integrar as TICs ao espaço urbano, tanto através das redes telemáticas e dos aparatos tecnológicos, quanto através de um espaço virtual [...], onde os cidadãos podem dispor de informações serviços e formas de interação”.

Souto, Dall’Antonia e Holanda (2006, p. 61), em uma tentativa de compreensão da conceituação de Cidade Digital, apontam características relacionadas ao fenômeno de digitalização dos serviços municipais, que não necessariamente se aduzem pela mera conversão daquilo que existe no mundo real para o virtual, assumindo um caráter de complementaridade entre ambos:

Por ser consequência de uma transformação em curso, a própria conceituação da cidade digital encerra complexidade e perspectivas distintas para definir sua essência e seus contornos. Compreender o alcance e as potencialidades do

processo de digitalização dos serviços e formas relacionais das comunidades em rede, sem necessariamente associá-lo a uma transposição do mundo real para o virtual, não é algo simples e de fácil apreensão, sobretudo se considerarmos que o espaço virtual pode ser, ele mesmo, o objeto da experiência. Com efeito, é muito mais plausível relacionar essa transição a uma complementaridade entre os dois mundos: o virtual e o concreto (Souto, Dall'Antonia and Holanda, 2006).

Na literatura que costuma tratar sobre Cidades Digitais, há a presença frequente de mecanismos de inovação tecnológica e comunicação no papel de revolucionar o prisma social atual, com a facilitação de acessos aos aparelhos públicos pelo ambiente virtual. Sob essa perspectiva, a Cidade Digital adquire um papel de inclusão social, aproximando o cidadão do Estado e de outras comunidades, seja a primeira relação pelo desenvolvimento de sistemas colaborativos e a segunda pela implementação de espaços virtuais que reduzem as barreiras geográficas, graças à implementação de uma infraestrutura de base eficiente (Andrade, Cruz and Rodrigues, 2020).

O Governo Federal, por intermédio do Ministério das Comunicações, instituiu um projeto para implantação das cidades digitais onde, desde já, outorga para si esta responsabilidade, restando as tarefas de gestão e manutenção para os municípios e Distrito Federal. O projeto possui 4 objetivos principais, que passam pela: 1) construção de redes digitais para comunicação nos municípios; 2) promoção e oferta de serviços digitais; 3) facilidade na apropriação das TICs pela gestão pública e pela população; e 4) construção de enlaces de conexão intermunicipais utilizando fibra óptica (Brasil, 2011).

Andrade, Cruz e Rodrigues (2021, p. 131) reforçam o entendimento anteriormente explorado sobre o projeto Cidades Digitais, complementando que estas devem tornar possível o desenvolvimento humano, econômico e cultural da sociedade, sempre com o auxílio de recursos de TIC:

Vê-se que o projeto “Cidades Digitais”, instituído pelo Governo Federal, simboliza o desígnio de promover investimentos para a construção de redes digitais que conectem os órgãos públicos, assim, proporcionando renovação da administração pública e ampliação do acesso à serviços governamentais (Andrade, Cruz and Rodrigues, 2021, p. 131).

Sobre o programa Cidades Digitais, pode-se afirmar que seu objetivo é fornecer uma infraestrutura de base aos municípios para que estes possam modernizar os serviços oferecidos à população. O referido programa foi dividido inicialmente em duas fases, em 2011 e 2013, onde os municípios interessados poderiam participar via edital. Além disso, foi realizada licitação para contratação das empresas que iriam implementar os sistemas de infraestrutura. Uma terceira fase foi iniciada em 2016 com o objetivo de transformar o programa Cidades Digitais em Cidades Inteligentes, integrando o projeto Brasil Inteligente e objetivando, além da ampliação da infraestrutura em fibra ótica, a instalação de centros de gestão de dados e redes de monitoramento e vigilância com câmeras (Anastácio, 2017, p. 8).

A partir desse contexto, é possível notar a evolução de um programa do Governo Federal que transforma um projeto de simples e estática digitalização dos serviços municipais para o alcance de um resultado mais autônomo, dinâmico e inteligente.

2.3. A cidade inteligente

O foco de uma Cidade Inteligente, é promover sustentabilidade com eficiência e planejamento consciente, resultando em bem-estar social, a partir da utilização de sistemas ecológicos que promovem qualidade de vida em um meio-ambiente saudável e equilibrado. O emprego de recursos de Tecnologia da Informação e Comunicação (TIC) também é um elemento considerável que pode atuar como combustível propulsor para o alcance desses objetivos. Em uma das concepções teóricas sobre esse assunto, a Cidade Inteligente é a integração conceitual entre Cidade Digital – aquela que utiliza os recursos de TIC na gestão municipal e atendimento ao público – e Cidade Sustentável – a que busca o bem-estar da sociedade (Guimarães, Araújo and Costa, 2020, p. 1090).

Reck e Vanin (2020, p. 466) seguem o mesmo pensamento sobre o emprego das TICs para a concepção de Cidades

Inteligentes, fazendo referência a outros estudos acerca do tema:

Os estudos que abordam o tema das *Smart Cities* tem sido recorrentes em afirmar que se trata de um conceito em construção, entretanto, já é possível observar um certo consenso, no sentido de que as ações com esta finalidade devem envolver o uso de tecnologia com enfoque nas pessoas e com o objetivo de proporcionar melhorias socioambientais e econômicas para as comunidades locais, abarcando um leque de soluções que transpassam os mais diferentes serviços públicos setoriais (Reck and Vanin, 2020, p. 466).

Weiss, Bernardes e Consoni (2017, pp. 4-5) também fazem referência a um arcabouço teórico que julga indiscutível a aplicação das diversas inovações tecnológicas ao conceito de Cidades Inteligentes, traduzindo-as naquelas cidades capazes de se conectarem à sociedade de maneira eficiente, melhorando as condições de sustentabilidade ou até fornecendo facilidades para desburocratização dos processos de gestão do município. Estas mesmas cidades possuem visão de futuro, baseando-se em decisões conscientes que englobam o acesso à saúde, educação, segurança pública, transportes e outras questões de infraestrutura com relevância social. Este autor ainda se refere à existência da Cidade Inteligente para fins de alavancagem de competitividade econômica e gestão ambiental.

Com a rápida e crescente urbanização, as cidades tornam-se lugares difíceis de se habitar. Em face dessa constatação, Moura (2021, pp. 31-33) enfatiza a importância da utilização das TICs como recurso útil na análise e otimização dos serviços e na adoção de políticas públicas, com a participação dos atores locais no desenvolvimento e implementação de processos e métodos que possam modernizar e fornecer um melhor atendimento à população, sempre prezando por uma boa gestão de recursos naturais e, conseqüentemente, melhoria na qualidade de vida.

Sob um ponto de vista alternativo, Porto Neto, Dornelles e Andrade (2021, p. 4) definem as Cidades Inteligentes como aquelas que “[...] conseguem fazer uso dos dados produzidos em massa pelos aparelhos das pessoas conectadas à internet

[...]” conseguindo, com isso, interpretar as informações inseridas para uma tomada de decisão mais adequada no sentido de disponibilizar os serviços públicos de maneira antecipada às necessidades da população. Os autores ressaltam a importância do desenvolvimento das *Smart Cities* coincidir com o avanço educacional, para evitar a clássica exclusão social onde apenas uma minoria possui acesso aos recursos tecnológicos.

Para o financiamento das Cidades Inteligentes não onerar excessivamente os recursos da administração pública, as parcerias público-privadas (PPPs) surgem como modelo de investimento nas TICs aplicadas à sustentabilidade. Desse modo, dispendiosas aplicações financeiras destinadas à implementação de tecnologias e comunicações em prol do bem-estar social deixam de sair dos cofres dos Municípios e da União, sendo aproveitadas em outras áreas de grande carência. Souza e Neto (2020) trazem alguns dos benefícios dessas parcerias:

Evidencie-se que diversos são os benefícios oriundos dessa opção pelas parcerias. A Lei nº. 11.079/2004 apresenta alguns, quais sejam: redução das despesas orçamentárias, profissionalização da gestão do serviço, implementação de projetos de grande vulto, possibilidade de remunerar o parceiro privado de acordo com seu desempenho, auxiliando, desta forma, no cumprimento das metas fiscais da Lei de Responsabilidade Fiscal (Souza and Neto, 2020).

3. Políticas de segurança pública em cidades inteligentes

Neste capítulo, apresentamos uma amostra do estado da arte sobre políticas públicas voltadas à Segurança Pública, no âmbito das Cidades Inteligentes.

Alguns exemplos de aplicação das TICs em políticas de Segurança Pública, de forma inteligente, foram constatados em países da América Latina, a exemplo da cidade de Medellín, na Colômbia, que implantou um centro de operações integrando todos os serviços de emergência e segurança, possibilitando que mais de 10 agências governamentais trabalhem de forma coordenada. A cidade de Medellin conta

com um número único de contato para esses serviços, centenas de câmeras de vigilância que apoiam o trabalho dos operadores de Segurança Pública e aplicativos para dispositivos móveis para denúncias anônimas com georreferenciamento (Moura, 2021, pp. 40-41).

No panorama Brasileiro, destacam-se os casos de São Paulo e Curitiba. A cidade de São Paulo possui como iniciativa de modelo para Cidades Inteligentes um projeto de compartilhamento de câmeras privadas com os órgãos de Segurança Pública (Polícia Militar e Civil), possibilitando também o monitoramento do tráfego nas vias públicas para detecção de congestionamentos, rastreamento de carros roubados, investigação de crimes e atendimento a ocorrências de enchentes, incêndios e acidentes de grande porte.

Além desse projeto, São Paulo também investiu em iluminação pública, substituindo luminárias por LED e fiação de cobre por alumínio, reduzindo o número de furtos desses cabos. Um programa de monitoramento da segurança municipal através da utilização de drones também se destaca dentre as políticas de Segurança Pública inteligentes (Moura, 2021, pp. 43-45).

Souza e Santos (2019, pp.148-149) em seu estudo, correlacionam o princípio da eficiência com a utilização de drones na segurança pública, onde observam que a tecnologia adotada apresenta resultados satisfatórios na prevenção e no combate a crimes:

A título de exemplo, o mais novo programa anunciado pela secretária de segurança pública do Estado de São Paulo, denominado de Dronopol, estabelece a inserção de drones em tempo real nas execuções das atividades de policiamento ostensivo e preventivo (Souza and Santos, 2019, pp. 149).

A cidade de Curitiba, considerada uma das cidades mais inteligentes do mundo, criou o Instituto das Cidades Inteligentes (ICI), responsável pelo desenvolvimento de soluções sustentáveis para a população. Como iniciativa de segurança pública, implantou um sistema de reconhecimento facial, facilitando o trabalho da Guarda Municipal e demais forças de segurança (Moura, 2021, pp. 48-50).

Moura (2021, p. 51), além dos casos supracitados, traz a experiência de Porto Alegre, que possui ações inteligentes em desenvolvimento, como o projeto de identificação automática de placas de veículos, com utilização das câmeras dos pardais, buscando rastrear pendências de pagamentos de impostos, furtos e roubos. Além disso, a cidade conta com um Centro In-

tegrado de Comando e Monitoramento, que fornece um canal de atendimento para todas as demandas de segurança.

4 Análise do cenário local

A primeira ação característica de uma Cidade Digital no município de Aracaju, noticiada no ano de 2010, faz referência à disponibilização de pontos de acesso sem fio com largura de banda de 1mbps, em lugares específicos, a exemplo da Orla de Atalaia, distribuída em quatro pontos, sendo um destes a Passarela do Caranguejo. O Parque Augusto Franco, mais conhecido como Parque da Sementeira também contou com os primeiros usos dessa tecnologia no âmbito do município. À época da sua implantação, planejava-se em uma segunda fase do projeto com a instalação no Bairro Bugio e, na terceira etapa, o alvo seria o Centro da Cidade, abrangendo as Praças General Valadão, Olímpio Campos e Fausto Cardoso. A quarta etapa seria direcionada ao calçadão da 13 de Julho (TV Cidade, 2010).

Partindo para o conceito de Cidade Inteligente, A Superintendência Municipal de Transportes e Trânsito (SMTT) de Aracaju estava com um projeto no ano de 2018, para ser concluído até o mês de maio de 2019, que visava a substituição de 150 semáforos tradicionais por equipamentos inteligentes. A vantagem do uso da tecnologia é a otimização do tráfego de veículos automotores, ciclistas e pedestres de maneira automática, com uma melhor gestão de tempo para abertura dos fluxos de passagem. Os benefícios do novo sistema seriam inclusivos aos deficientes visuais, onde as botoeiras de passagem possuiriam sinais sonoros de orientação, além do alto relevo para leitura em braile (TV Atalaia, 2018).

Seguindo o mesmo modelo de outras localidades, Aracaju possui um Centro de Integrado de Operações em Segurança Pública (CIOSP) desde o ano de 2009, com o objetivo de concentrar chamadas voltadas aos serviços de emergência pelo número 190. Em adição à sua função primária de atuação, o CIOSP funciona como serviço de monitoramento por câmeras de vigilância, otimizando os atendimentos a crimes de roubo, furto e dano ao patrimônio (Prefeitura de Aracaju, 2009).

É indubitável a necessidade de existir toda uma infraestrutura de rede por trás de qualquer iniciativa de Cidades Digitais ou Inteligentes. Nesse sentido, durante o período pandêmico, onde houve a necessidade de atender os novos requisitos de ensino à distância, o município de Aracaju firmou um convênio com o Tribunal de Justiça de Sergipe (TJ-SE) para utilização de seu anel óptico com 62km de extensão. Além da ampliação territorial da rede municipal, ganhou-se também em largura de banda. Onde operava-se em 10mpbs, passou a haver conectividade gigabit, cerca de cem vezes mais rápida que a anterior, resultando em uma banda de 1 gbps (Prefeitura de Aracaju, 2019).

Outro recurso bastante utilizado na Segurança Pública, é a consulta às bases de dados alimentadas por Boletins de Ocorrência, que possibilitam um trabalho de inteligência para mapeamento de manchas criminais. A utilização dessa tecnologia permite a destinação dos recursos de segurança (investigações, polícia ostensiva, viaturas, etc.) de forma eficiente, tendo como resultado a melhoria nos índices de criminalidade (SSP-SE, 2020).

Todas as ações anteriormente citadas são projetos inteligentes que atuam diretamente ou indiretamente para a melhoria da Segurança Pública. O Ranking Connected Smart Cities classifica o município de Aracaju na posição nº 60 em relação aos demais municípios brasileiros, quando o indicador abordado é a Segurança Pública. As variáveis utilizadas por esse ranking são: 1) Existência de um centro de controle e operações; 2) Despesas per capita com segurança; 3) Homicídios / 100 mil habitantes; 4) Mortes em acidentes de trânsito / 100 mil habitantes; 5) Policiais, guardas-civis municipais e agentes de trânsito por 100 mil habitantes; e 6) Monitoramento de área de risco. Existem outros rankings para mensurar Cidades Inteligentes e algumas de suas variáveis podem ou não coincidir (Urban Systems, 2022).

5. Considerações finais

Modernizar os ambientes urbanos que demonstram um crescimento acelerado e em muitas ocasiões desordena-

do, reflete o necessário emprego das Tecnologias da Informação e Comunicação para facilitar os processos de gestão e atendimento à população, fenômeno relacionado ao conceito de Cidades Digitais. Este raciocínio deve sempre ser acompanhado da ideia de se manter uma sociedade sustentável, com consumo consciente e reaproveitamento adequado dos recursos naturais. As Cidades Inteligentes mostram-se como conceitualização voltada à solução dos problemas do crescimento e a preocupação com o futuro da sociedade, já que esta depende do equilíbrio ambiental para o seu desenvolvimento.

Durante a implantação do Programa Cidades Sustentáveis no Brasil, percebe-se a preocupação em adequá-lo aos Objetivos de Desenvolvimento Sustentável promovido pela Organização das Nações Unidas, adotando um corpo de maior responsabilidade e seriedade no cumprimento e na realização de suas metas. A Cidade Inteligente, antes de tudo, é uma Cidade Sustentável, seja com ou sem o amparo das Tecnologias de Informação e Comunicação.

A importância desta pesquisa evidencia-se na demonstração das iniciativas relacionadas às Cidades Inteligentes na capital do estado de Sergipe. Constata-se que, apesar dos investimentos alocados pela União, para tal finalidade, existe uma inércia na implementação destas iniciativas, além de uma subutilização dos recursos tecnológicos de infraestrutura já disponibilizados. O evento pandêmico mostrou-se como um acelerador local da expansão dessas tecnologias, perante a necessidade da continuidade social em um momento de extremo confinamento, na utilização de serviços básicos como saúde, educação e segurança pública.

Como trabalhos futuros, sugere-se o acompanhamento da efetividade das ações que foram implantadas pelo município de Aracaju, acerca das Cidades Inteligentes no escopo da Segurança Pública e em outras dimensões como Saúde, Educação, Transportes e qualquer outra tarefa de responsabilidade da administração direta e indireta que possa afetar o bem-estar da população, além de mapear novos projetos que venham surgir nesse sentido.

Referências

- Anastácio, K. (2017). 'Difusão de Políticas Públicas & Tecnologia: expansão das Cidades Digitais no Brasil', *9º Congresso Latino americano de Ciência Política*, Montevideu, pp. 1-23.
- Andrade, D. C. M, Cruz, L. F. S. and Rodrigues, F. F. (2021) 'Cidades brasileiras digitais: mapeamento de projetos e implementações a partir de políticas públicas de iniciativas tecnológicas e seu subsídio ao desenvolvimento'. *Cardenos de Graduação UNIT*, 6(3), pp. 129-142 [online]. Available at: <https://periodicos.set.edu.br/cadernohumanas/article/view/9986> (Accessed: 14 August 2023).
- Andrade, D. C. M, Cruz, L. F. S. and Rodrigues, F. F. (2020) 'Cidades Digitais e Sociedade em Rede: Interseções e desafios de uma construção sociotécnica', *Interfaces Científicas*, 10(2), pp. 66-79.
- Andrade, D. C. M. and Cruz, L. F. S. (2019) 'Do Espaço Ao Ciberespaço: Cidades Digitais e a Sociedade (In)Formacional em Questão', *VIII CONINTER – Congresso Internacional Interdisciplinar em Sociais e Humanidades*, pp. 1-10.
- Andrade, D. C. M. and Meira, M. D. (2021) 'Cidades digitais e sua compatibilidade com a segurança pública para a efetivação dos direitos humanos', *Revista Em Tempo*, 21, pp. 68-82 [online]. Available at: <https://revista.univem.edu.br/emtempo/article/view/3291> (Accessed: 14 August 2023).
- Botton, G. Z. et al (2021) 'As Construções das Abordagens Conceituais de Cidades Sustentáveis e Inteligentes para Superar os Desafios dos Objetivos do Desenvolvimento Sustentável', *Desafio Online*, 9(3), pp. 619-642 [online]. Available at: <https://desafioonline.ufms.br/index.php/deson/article/view/13072/9369> (Accessed: 14 August 2023).
- Brasil. *Portaria nº 376, de 19 de agosto de 2011* [online]. Available at: https://www.gov.br/mcom/pt-br/aceso-a-informacao/legislacao/copy_of_PORTARIAN376DE19DEAGOSTO-DE2011.pdf. (Accessed: 14 August 2023).

- Fabris, J. et al (2020) 'Cidades Sustentáveis: Caminhos e Possibilidades', *International Journal of Professional Business Review*, São Paulo, 5(2), p. 214-233 [online]. Available at: <https://dialnet.unirioja.es/servlet/articulo?codigo=7593371>. (Accessed: 14 August 2023).
- Guimarães, P. B. V., Araújo, D. S. and Costa, A. A. (2020), 'A implantação de cidades inteligentes no nordeste brasileiro: um breve Diagnóstico', *Revista de Direito da Cidade*, 12(2). p. 1084-1104. Available at: <https://www.e-publicacoes.uerj.br/index.php/rdc/article/view/39957>. (Accessed: 14 August 2023).
- Lemos, A. (2007) 'Cidades Digitais: portais, inclusão e redes no Brasil'. Salvador: EDUFBA.
- Moura, G. S. (2021) *Cidades Inteligentes: Alternativas para a melhoria da segurança pública em Porto Alegre*. Porto Alegre: UFRGS [online]. Available at: <https://www.lume.ufrgs.br/bitstream/handle/10183/248519/001145729.pdf> (Accessed: 14 August 2023).
- Porto Neto, A. A., Dornelles, C. J. H. C. and Andrade, D. C. M. (2021) 'O Direito da Personalidade à Educação, as Novas Tecnologias de Ensino no Brasil e seus Reflexos para as Cidades Inteligentes', *Anais Do Simeeduc*, pp. 1-15.
- Prefeitura de Aracaju (2009) *Edvaldo Nogueira acompanha inauguração do CIOSP 190* [online]. Available at: <https://www.aracaju.se.gov.br/noticias/37528/> (Accessed: 14 August 2023).
- Prefeitura de Aracaju (2019) *Rede de fibra ótica irá proporcionar internet mais veloz nas escolas do Município* [online]. Available at: <https://www.aracaju.se.gov.br/noticias/81436/> (Accessed: 14 August 2023).
- Programa Cidades Sustentáveis (2016) *Guia de Gestão Pública Sustentável – atualizado com os ODS. Versão resumida*. São Paulo [online]. Available at: https://www.cidadessustentaveis.org.br/arquivos/publicacoes/gps_guia_gestao_publica_sustentavel.pdf (Accessed: 14 August 2023).

- Reck, J. R. and Vanin, F. S. (2020) 'O Direito e as Cidades Inteligentes: desafios e possibilidades na construção de Políticas Públicas de Planejamento, Gestão e Disciplina Urbanística', *Revista De Direito Da Cidade*, 12(1), pp. 464-492 [online]. Available at: <https://www.e-publicacoes.uerj.br/index.php/rdc/article/view/39618/33502> (Accessed: 14 August 2023).
- Secretaria de Segurança Pública de Sergipe (2020) *Boletim de ocorrência identifica mancha criminal e orienta operações de combate à criminalidade. Secretaria da Segurança Pública* [online]. Available at: <https://www.ssp.se.gov.br/Noticias/DetaIhes?idNoticia=16681> (Accessed: 14 August 2023).
- Souto, Á. A., Dall'antonia, J. C. and Holanda, G. M. de. (2006) *As Cidades Digitais no mapa do Brasil: Uma rota para a inclusão social*. Brasília: Ministério das Comunicações.
- Souza, P. V. N. C. S. and Neto, R. S. (2020) 'Perspectivas das Cidades Inteligentes na Administração Pública em Tempos de Transformação Digital', *Revista Jurídica – UNICURITIBA*, Curitiba, 3(60), pp. 39-68 [online]. Available at: <http://revista.unicuritiba.edu.br/index.php/RevJur/article/view/4172/371372499> (Accessed: 14 August 2023).
- Souza, P. V. N. C. and Santos, A. T. (2019) 'A Inserção dos Drones (RPAs) Na Segurança Pública Brasileira: uma análise sob a ótica do princípio da eficiência', *Revista Em Tempo – UNIVEM*, 18(1), pp. 133-155 [online]. Available at: <https://revista.univem.edu.br/emtempo/article/view/3209> (Accessed: 14 August 2023).
- TV Atalaia (2022) 'Foragido por tráfico de drogas é preso no Pré-Caju após leitura facial de câmeras', *Cidade Alerta. YouTube*, 7 November [online]. Available at: <https://www.youtube.com/watch?v=xzjWB9Re2-M> (Accessed: 14 August 2023).
- TV Atalaia (2018) 'SMTT inicia instalação dos semáforos inteligentes em Aracaju' *Cidade Alerta. YouTube*, 16 November [online]. Available at: https://www.youtube.com/watch?v=u_TbR2OegR8 (Accessed: 14 August 2023).

- TV Cidade (2010) *Wi-fi em Aracaju*. YouTube, 15 December [online]. Available at: https://www.youtube.com/watch?v=PLhMN7z0j_E. (Accessed: 14 August 2023).
- Urban Systems (2022) *Ranking Connected Smart Cities 2022*. São Paulo [online]. Available at: <https://ranking.connectedsmartcities.com.br> (Accessed: 14 August 2023).
- Weiss, M. C., Bernardes, R. C. and Consoni, F. L. (2017) 'Cidades inteligentes: casos e perspectivas para as cidades brasileiras', *Revista tecnológica*, 5(1), pp.1-18.

MEIO AMBIENTE DIGITAL, INTELIGÊNCIA ARTIFICIAL E EDUCAÇÃO

Alexandre Cortez Fernandes¹

Patrícia Montemezzo²

DOI: 10.29327/5312711.1-11

Sumário: 1. Introdução. 2. Verificações acerca da Inteligência Artificial. 3. Possibilidades da Inteligência Artificial. 4. Riscos da virtualidade e a questão da educação. 5. Considerações finais. Referências.

1. Introdução

No atual estado da vida contemporânea, parece ser inconcebível a vida sem as mídias digitais. Importa aclarar que a expressão *digital* é compreendida em oposição às mídias analógicas, que caracterizam os meios de comunicação social.

Neste sentido, fala-se atualmente em meio ambiente digital, considerando a ampliação do conceito para abranger também bens jurídicos criados a partir da atual era da hiperconectividade. Da mesma forma, a inteligência artificial hoje se faz presente no cotidiano, nos mais diversos dispositivos e aplicativos, integrando intimamente o meio ambiente digital.

Não há dúvidas quanto aos desafios que vida contemporânea, inserida no universo das mídias digitais, oferece. Desafia, inclusive, educadores, que precisam formar cidadãos capazes de portar-se e agir no meio ambiente digital, tirando dele o melhor proveito possível. Não basta saber operar as ferramentas dotadas de inteligência artificial, mas também posicionar-se diante das interferências que podem decorrer delas.

1 Doutor em educação, professor da graduação e do PPG em Direito da Universidade de Caxias do Sul (UCS). E-mail: acfernand@ucs.br.

2 Mestre em direito, professora de Direito da Universidade de Caxias do Sul (UCS) e advogada. E-mail: pmonteme@ucs.br.

Em razão disso tudo é que o presente artigo visa analisar a educação ambiental como instrumento de formação para utilização humanista da inteligência ambiental, no ambiente da hiperconectividade. Análise de extrema relevância, considerando-se o avanço crescente de novas tecnologias dotadas de capacidade emulatória do pensamento humano.

Pode-se refletir, ainda, sobre uma evolução da tecnologia tal, a ponto de que o homem venha a ser substituído pela máquina, no trabalho. Ou ainda, que essa o influencie na tomada de decisão sob diversos aspectos, tanto na vida profissional quanto na pessoal, e, também, no exercício da cidadania e convívio social. Diante disso é que o debate acerca do papel da educação ambiental mostra-se fundamental, a fim de assegurar o protagonismo do homem, mesmo diante de uma tecnologia extremamente influente e presente.

Os objetivos específicos do presente estudo são os seguintes: perceber se é possível utilizar-se das concepções do direito ambiental no ambiente da hiperconectividade; demonstrar algumas das possibilidades da inteligência artificial na contemporaneidade; apontar alguns riscos do excesso de exposição ao ambiente virtual; e indicar a educação ambiental como uma possível resposta à crise do excesso de exposição.

A metodologia utilizada no artigo está pautada pelo tipo qualitativo e a técnica usada foi a bibliográfica-exploratória.

O texto está dividido da seguinte maneira: inicialmente, tecem-se algumas ponderações sobre a hiperconectividade e o meio ambiente digital. Num segundo momento, há raciocínios sobre as possibilidades de inteligência artificial. O trabalho culmina com o apontamento de alguns riscos sobre essa prática, em especial, sob o ângulo da educação. Essa educação, frise-se, é a educação ambiental, que deve visar a melhor formação do educando em vistas à cidadania digital.

2. Verificações acerca da Inteligência Artificial

A intensificação do uso da tecnologia que se propõe a conectar pessoas e dispositivos, caracteriza a chamada hiper-

conectividade. Essa forma de comunicação liga pessoas de modo instantâneo em toda parte do mundo, disseminando de uma forma global as informações. Essa estrutura é o reflexo de uma nova maneira de se comunicar, de novas formas de produzir, aprender, trabalhar, interagir socialmente, refletir, não deixando de ser uma configuração de meio ambiente.

Há autores que compreendem essa conexão como uma espécie de reconfiguração da interação entre as pessoas num meio ambiente digital. Fiorillo e Souza (2016) indicam que não é equivocado considerar o meio ambiente digital como uma ramificação do direito ambiental. Nesse sentido, estudaria os bens portadores de referências às ações, memória, à identidade dos grupos e sistemas que fazem parte da sociedade brasileira no meio digital, sendo certo que, também congregaria a rede mundial de computadores.

Fiorillo e Souza (2016) consideram que esse possível novo ramo do Direito Ambiental teria por objetivo identificar e buscar regular o uso dos bens jurídicos que podem ser afetados pela hiperconectividade atual. Cavedon et al. (2015) defendem que a proteção constitucional ao meio ambiente é mais extensa e transcende a concepção ecológica, abrangendo uma série de outros dispositivos que, direta ou indiretamente, se relacionam a valores ambientais. Trata-se de uma nova categoria de proteção ao meio ambiente, que se baseia no reconhecimento constitucional de que a cultura e a tecnologia também acabam fazendo parte do patrimônio cultural brasileiro. Ao se aceitar essa ponderação, o meio ambiente digital é considerado como sendo parte integrante e inseparável do conceito jurídico de meio ambiente, estabelecido pela Lei n. 6.938/81, sendo percebido também por intermédio de sua inequívoca dimensão cultural. (Cavedon et al, 2015).

Nessa esteira, não parece ser uma novidade a definição de meio ambiente que insere a cultura como um componente que o integra. Basta indicar as dimensões plurais que envolvem a proteção ambiental prevista na Constituição Federal, assim como na Lei n. 6.938/81, abrangendo o meio ambiente urbano, o meio ambiente do trabalho, além do ecológico e do cultural. Assim, dada a concomitância, a influência e a im-

portância da hiperconectividade na sociedade atual, pode-se pensar em afirmar a existência de um meio ambiente digital.

Importa verificar que não é propriamente pela hiperconectividade que haveria um meio ambiente digital. Mas, na esteira do art. 3º, da Lei n. 6.938/81, que dispõe sobre a Política Nacional do Meio Ambiente, indica, *in verbis*:

Art 3º – Para os fins previstos nesta Lei, entende-se por:
I – meio ambiente, o conjunto de condições, leis, influências e interações de ordem física, química e biológica, que permite, abriga e rege a vida em todas as suas formas;

Cumprе salientar que o ambiente digital pode ser considerado um meio ambiente. O ambiente digital é um locus não só de experimentações, mas, também, de compartilhamentos. Por ocasião de sua disseminação universal pareceu ser a quebra da utopia da Liberdade, pela sua ausência de limites e sem demarcações de fronteiras. A humanidade experimentava a certeza de que as suas inovações seriam a tônica do desenvolvimento planetário.

O ambiente digital é, portanto, um meio em que as relações sociais, de trabalho, de estudo e até afetivas acontecem. Está cada vez mais presente em todos os aspectos da vida contemporânea, a ponto de tornar-se impossível a vida sem a utilização dos recursos tecnológicos. Diante disso é que parece impossível dissociar da concepção de meio ambiente o aspecto digital.

Ademais, o termo ecossistema foi cooptado pela linguagem da inovação. Para a biologia, ecossistema é um conjunto de interações entre os organismos vivos e os componentes não vivos do ambiente. Para a linguagem da inovação, que é uma linguagem de mercado, considera-se um ecossistema de inovação a junção, em ambiente colaborativo, de Poder Público, instituições de ensino superior e de empresas e organismos da iniciativa privada que operam em conjunto, para a obtenção de um resultado em comum.

Nessa linha de identificação de ecossistemas, portanto, parece legítimo se pensar num ecossistema digital, que pode ser considerado o conjunto de aplicativos, plataformas, sites e ferramentas que, de modo unificado e orgânico, buscam alcan-

çar uma melhor colocação de um serviço, de um portal ou de uma identificação específica. Assim, parece ser um ecossistema a comunidade no mundo digital um grupo de pessoas que interagem no meio virtual, em compartilhamento de recursos.

Em sendo assim, mostra-se possível retomar a concepção do art. 3º, da Lei n. 6.938/81, no que tange ao mundo virtual, com uma aplicação desde que logicamente possível, dos princípios de direito ambiental, constitucionalmente derivados, sendo um deles, a adequada possibilidade de uma educação ambiental.

3. Possibilidades da Inteligência Artificial

A era da hiperconectividade, impulsionada pela disseminação da internet e das redes sociais, por gerar grande volume de dados decorrentes das conexões entre pessoas e dispositivos, necessita cada vez mais de mecanismos tecnológicos. É nesse contexto que a inteligência artificial se torna cada vez mais relevante, pois permite a análise e o processamento de grandes quantidades de informações em tempo real. Com o avanço da inteligência artificial, surgem novas possibilidades para a automação de tarefas e a tomada de decisão baseada em dados, tornando-se um recurso valioso para diversos setores.

No esforço de construção de uma conceituação para a inteligência artificial, é importante lembrar as origens da estruturação de sistemas em redes neurais, arquitetura responsável pela alta capacidade de autonomia de sistemas inteligentes.

Essa importantíssima estruturação ocorreu na década de 1980, quando um grupo de pesquisadores propôs as redes neurais para aprendizado de máquina e ganhou reconhecimento em 2012 ao vencer a competição ImageNet. Desde então, as redes neurais se tornaram onipresentes, recebendo investimentos das gigantes de tecnologia, como o Google. As redes neurais, por meio de algoritmos de inteligência artificial, são capazes de estimar a probabilidade de eventos ocorrerem, como identificar um tumor maligno num corpo, a presença de um animal numa imagem, o candidato apropriado para deter-

minada função, o tipo de serviço ou de produto adequado aos desejos de um determinado consumidor (Kaufman, 2022).

Apesar disso, conforme defende Kaufman (2022), esses sistemas ainda carecem da essência da inteligência humana, como a capacidade de compreender o significado e fazer analogias e generalizações. Portanto, conforme o autor, ainda há pouco progresso em prover a inteligência artificial de senso intuitivo, de forma que, apesar da grande acurácia dos sistemas, hoje, não se verifica algum que se assemelhe, de fato, ao cérebro humano, que é capaz de processar informações em diferentes contextos, e associando aspectos não meramente lógicos, como sentimentos, por exemplo. Um sistema inteligente não é dotado de senso intuitivo, tampouco complexo como o pensamento do homem.

Ainda no esforço de melhor compreender o conceito e a amplitude dos sistemas inteligentes, Freitas (2020) diferencia simples automação e inteligência artificial, já que a capacidade de aprendizagem é uma das principais características desta tecnologia. A automação não inclui esta capacidade, pois trata-se de processo estritamente mecânico, ao passo que a inteligência artificial abrange aspectos que a aproximam da inteligência humana, sendo composta por um conjunto de algoritmos programados para cumprir objetivos específicos. O autor afirma que “a inteligência artificial é um sistema algorítmico adaptável, relativamente autônomo, emulatório da decisão humana” (Freitas and Freitas, 2020, p. 30).

Sistemas inteligentes têm, portanto, a capacidade de evoluir e se modificar conforme as informações são processadas, aprendendo por si, através de técnicas de *machine learning*. Uma das modalidades mais avançadas desse tipo de aprendizado de máquina se dá através de *deep learning*, que se vale de estruturas semelhantes às redes neurais humanas, contribuindo para algoritmos de inteligência artificial extremamente avançados e capazes de aprender com muita facilidade, processando informações de forma muito eficiente. (Freitas and Freitas, 2020).

O aprendizado de máquina decorrente de estruturas de *machine learning* e *deep learning* fazem com que os sistemas

inteligentes tenham autonomia no processamento de dados e informações, o que caracteriza um algoritmo de inteligência artificial, juntamente com a capacidade de compreender a linguagem natural (humana). É especialmente essa autonomia na aprendizagem que assemelha um sistema algorítmico à inteligência humana, ao pensamento humano – insista-se que é uma similitude, nunca uma substituição na plenitude.

Lacerda (2022) indica que a inteligência artificial não é uma cópia do pensar humano, mas um modelo do pensamento baseado na lógica – vê-se daí, portanto, que os limites da lógica passam a ser os limites da inteligência artificial. Nessa linha, é possível compreender que sistemas inteligentes operam através de métodos estatísticos que analisam e processam grandes volumes de informação.

Desse modo, não há como deixar de associar a inteligência artificial com a aprendizagem humana. Essa relação também foi apresentada por Turing (1950), que propôs um conceito amplamente utilizado atualmente sobre a inteligência artificial. Para o autor, uma máquina só pode ser considerada inteligente se apresentar duas características fundamentais que se interconectam: aprendizado e autoaperfeiçoamento, com o objetivo de se aproximar, ou pelo menos tentar reproduzir, a mente humana (Turing, 1950). O autor compara o aprendizado da máquina ao aprendizado de uma criança, dividindo-o em três fases: o estado inicial da mente da máquina, equivalente ao nascimento da criança; a educação recebida pela criança, representada pelos dados programados na máquina; e a experiência adquirida pela vivência, não necessariamente educacional (Turing, 1950). Inclusive, a partir desta concepção é que se criou o conceito de máquina criança. Muito embora possam ser tecidas críticas à acepção não tão humanista da comparação, a reflexão parece válida para perceber-se as características interconectadas.

Turing (1950) também criou uma testagem para determinar a capacidade da inteligência artificial de alcançar um nível satisfatório. Trata-se do denominado Jogo da Imitação, que consiste em uma simulação que envolve três participantes: uma pessoa, um computador e um interrogador humano,

que ficam em salas separadas e só podem se comunicar por texto datilografado. O objetivo do teste é ludibriar o juiz, fazendo com que a máquina seja capaz de se passar por um ser humano na conversa. O juiz deve analisar o conteúdo da conversa e tentar distinguir qual é a máquina e qual é o ser humano. Essa forma de teste ficou conhecido como o Teste de Turing (Onody, 2021). Percebe-se que o teste almeja verificar a possibilidade de uma máquina comportar-se de modo equivalente a um ser humano ou de modo distinto a uma pessoa.

Com o avanço e sofisticação da inteligência artificial, parece provável que, brevemente, nas experiências diárias de interação, o homem tenha dúvidas se está interagindo com pessoas humanas ou com sistemas inteligentes. Inclusive, vive-se uma amostra dessa situação. Atualmente, por exemplo, o uso de assistentes pessoais como Alexa e Siri, oportunizam uma conversa, muito embora bastante limitada em termos dialógicos. Ainda assim, esses assistentes são capazes de compreender e atender a comandos de voz, apresentando uma resposta com grande grau de assertividade, diante da acurácia do método estatístico que operam.

Também já existem sistemas inteligentes programados para geração de texto de forma muito otimizada, como é o caso do ChatGPT, produto da empresa OpenAI – sendo certo que também apresentam limitações reflexivas e analíticas, mas, ainda assim, são textos com muitos elementos coerentes. Considerando a importância da linguagem, esses exemplos de inteligência artificial surpreendem pela agilidade e assertividade com que as informações são processadas e respostas, verbais ou por escrito, são fornecidas.

Lacerda (2022), buscando uma classificação para a inteligência artificial no Direito, defende que a inteligência artificial é adequada à categoria dos bens jurídicos imateriais, já que é composta por elementos imateriais de software. Além disso, a inteligência artificial trabalha com dados, informações e processamentos destes, com o objetivo de alcançar uma finalidade específica. Entretanto, é importante a análise quanto à necessidade de elementos materiais, como hardwares, robôs e equipamentos eletrônicos em geral. Diante disso, a inteligência artifi-

cial se aproximaria da categoria de bens móveis, fungíveis ou infungíveis (Lacerda, 2022). Essa determinação é de relevo, por todos os elementos jurídicos que lhes são decorrentes, notadamente no que diz respeito aos negócios jurídicos.

Qualquer que seja a classificação de bem jurídico que se atribua à inteligência artificial, o certo é que interfere e participa cada vez mais da vida humana, a ponto de se tornar impossível pensar em se afastar dela. É, portanto, objeto de análise e regulação jurídica, visando definir seus limites, repercussões e até responsabilidades. Tal temática afasta-se dos objetivos da presente análise, razão pela qual não será aprofundada, fazendo-se necessário apenas salientar que ainda não existe marco legal específico sobre o tema, embora alguns Projetos de Lei estejam em tramitação e até estágio avançado nas Casas Legislativas Brasileiras, o que demonstra a efetiva necessidade de uma regulação.

Sistemas inteligentes estão presentes em diversas ferramentas no cotidiano, desde utensílios domésticos, como, por exemplo, máquinas de lavar, aspiradores de pó e geladeiras; até sistemas de gestão em empresas, alguns deles capazes de auxiliar na tomada de decisão na seleção de empregados ou na identificação de riscos para um empreendimento, apenas para exemplificar. Crianças utilizam sistemas inteligentes em dispositivos e softwares de educação e entretenimento, através de robôs e ferramentas de programação. O Poder Judiciário já está instrumentalizado, uma vez que diversos Tribunais possuem softwares que automatizam processos de forma autônoma. Assim também acontece em diferentes esferas da Administração Pública brasileira que utilizam sistemas de gestão inteligentes em relação ao Sistema Único de Saúde, na área da segurança pública, como auxiliar no controle de tráfego urbano, dentre outros. Isso tudo, sem considerar o modelo eleitoral, o sistema de recebimento de declaração de imposto de renda, ambos com um acesso massificado.

Neste contexto, crescem os debates acerca da possibilidade de substituição do homem pela inteligência artificial, principalmente pela eficiência e assertividade dos sistemas, em especial na automatização de processos burocráticos, que,

quando realizados manualmente, demandam grande tempo e esforço.

Sistemas inteligentes executam essas atividades burocráticas e repetitivas com extrema facilidade, gerando uma perigosa ideia de que, em razão dessa facilidade e imediatidade, a pessoa humana poderia ser descartada dessas práticas. Essa problemática fica ainda mais reforçada pela questão de que a inteligência artificial, através de mecanismos de *deep learning*, tem potencial ainda maior, sendo capaz de tomar decisões, de fazer escolhas igualmente com grande eficiência e assertividade.

Entretanto, por outro lado, defende-se que sistemas inteligentes são ferramentas que trarão cada vez mais facilidades para o homem, e que têm justamente a finalidade de auxiliá-lo, não de substituí-lo. Por mais eficiente ou inteligente que possa ser uma máquina, esta jamais será dotado de características humanas. Perceba-se que essa concepção é cercada de polêmicas e que não se pode, ingenuamente, crer que essa substituição do ser humano no processo produtivo é somente para o benefício da pessoa – há muita percepção de auferimento de lucro nessa troca.

Rech (2020) toma por argumento a teoria evolutiva de Darwin e indica que a inteligência artificial pode ser um resultado natural da evolução do homem, estando até mesmo inserida no conceito de meio ambiente:

No contexto da discussão sobre inteligência artificial, partimos da seguinte afirmativa filosófica: não existe vida artificial, assim como não existe inteligência artificial, porque se é artificial não é vida nem inteligência. A vida inteligente é filha legítima da natureza. Searle afirma que a máquina não entende o significado do que faz, pois não tem consciência, pois a consciência e, com ela, o significado, só podem ser produzidos por cérebros vivos e nunca por uma máquina. Essa afirmativa reforça a provocação filosófica inicial de que a inteligência é natureza, vida, em potência. Seria necessário criar um cérebro humano e replicá-lo na máquina, para podermos falar em inteligência artificial. Mas, se a tecnologia conseguir fazer isso, estaremos definitivamente afirmando que a origem da inteligência é a natureza, pois o cérebro humano nada mais é do que a evolução natural das espécies (Rech, 2020, pp. 24-25).

Efetivamente, há entendimentos no sentido de que a inteligência artificial, através da biotecnologia, poderá avançar de forma a ser capaz de reproduzir até mesmo reações físicas, fisiológicas humanas, como defende Harari (2018). Embora o autor sustente a ideia de que se está distante disso, pelo estado atual da inteligência artificial, não descarta a possibilidade de criação de seres humanos sintéticos, o que desafiaria sobremaneira a ética (Harari, 2018). A partir dessa verificação do autor, inclusive, haveria não um desafio, mas, talvez, uma contrariedade enorme à ética, esta sim, a desafiar a educação ambiental.

Mesmo que se considere que sistemas inteligentes podem desempenhar tarefas complexas, um algoritmo é projetado para realizar tarefas específicas e não pode ter a mesma flexibilidade e criatividade do ser humano. Não terá capacidade de improvisar ou adaptar-se a imprevistos. Também não será dotada de empatia, não conseguindo entender completamente ou sentir emoções humanas – o que é importante em diversas áreas, inclusive a educação e o Direito.

Admitir a possibilidade de que a máquina possa pensar ou criar pensamento, conhecimento, inteligência é o mesmo que aceitar a possibilidade de a criatura, ou a forma material, ter potência de criar o criador. Isso não tem lógica tampouco racionalidade, pois seria inverter a própria ordem da natureza ou admitir que a máquina se constitui fase avançada da evolução das espécies preconizada por Darwin (Rech, 2020, p. 36).

Assim, a inteligência artificial não poderá tomar decisões éticas ou morais, já que envolvem questões subjetivas que ultrapassam o alcance do aprendizado de máquina. Por mais presentes que se tornem, os sistemas inteligentes não deverão alcançar e substituir o cérebro humano, especialmente se considerar-se que um decorre do outro; ou seja, sem a natureza humana, não se tem inteligência artificial.

Entretanto, os desafios enfrentados em vários segmentos, diante dos avanços da inteligência artificial são latentes, o que não é diferente na área da educação.

4. Riscos da virtualidade e a questão da educação

Conforme analisado, inúmeras são as possibilidades da inteligência artificial, diante da sua autonomia, tanto no funcionamento, quanto na aprendizagem (aprendizagem de máquina). Entretanto, há riscos no uso de inteligência artificial que sequer podem ser adequadamente mensurados, mas que devem ser debatidos até mesmo em vistas à melhor regulação jurídica.

Doneda (2018) sintetiza parte desses riscos quando afirma que a inteligência artificial pode prever o comportamento individual, o que pode impactar diretamente os direitos individuais, incluindo a autonomia, a igualdade, a liberdade e a personalidade.

Na sociedade atual, os dados pessoais são a única forma de representação das pessoas perante organizações estatais e privadas, e decisões automatizadas baseadas em métodos estatísticos podem abrir ou fechar oportunidades e acessos. Se houver uma representação equivocada em determinados contextos sociais, pode afetar a integridade moral e a personalidade do indivíduo, bem como limitar a sua liberdade de ação e escolhas existenciais – nesse particular verificam-se lesões a direito da personalidade, o que, em termos jurídicos ensejam indenização por dano moral.

Além disso, a classificação e seleção operadas por algoritmos podem produzir resultados desiguais para pessoas em situações semelhantes, violando o princípio da igualdade (Doneda, 2018). Trata-se de uma real possibilidade de discriminação realizada por algoritmos, que se torna possível em razão de que estes não criam informações, mas geram resultados a partir de dados a eles fornecidos.

Frazão (2021) defende que problemas de discriminação estatística ocorrem mesmo quando os dados e as estatísticas estão corretos. Quanto mais enraizado um preconceito na vida real, mais os algoritmos têm a tendência de identificá-lo como um padrão e replicá-lo, já que opera mediante métodos estatísticos. Esses problemas se agravam quando os dados não são de qualidade ou quando ocorrem falhas na metodologia esta-

tística ou na interpretação dos resultados. A qualidade dos dados utilizados pelos sistemas algorítmicos nem sempre é assegurada, especialmente em relação a dados que podem mudar rapidamente, como preferências ou hábitos individuais. A falta de mecanismos para desconsiderar dados ultrapassados e utilizar apenas dados atuais e confiáveis pode resultar em ruídos poderosos que mantêm ou incluem indivíduos em determinados grupos com base em critérios que não são mais relevantes no presente.

Ser substituído pela inteligência artificial não parece ser a real preocupação que o homem deve alimentar, e, sim, o risco real de que o algoritmo pratique algum tipo de discriminação. Além disso, há o receio de que ela venha a influenciar sobremaneira sua vida, suas decisões, seus direitos e até seu livre arbítrio. Isso porque os algoritmos, especialmente de redes sociais e outras ferramentas de comunicação, que também são dotados de inteligência artificial, são programados para estimular a permanência das pessoas na conexão, direcionando conteúdos de acordo com as preferências geradas a partir de perfis gerados através de grandes bancos de dados (*Big Data*).

Assim, automaticamente inúmeros conteúdos são direcionados, de acordo com o engajamento que podem vir a provocar, observando os gostos pessoais e as características de consumo de cada usuário. Esse direcionamento estimula a dependência cada vez maior desses meios de comunicação, e acaba por alavancar comportamentos nocivos, agressivos e de ódio (Fisher, 2023).

Kaufman (2022) manifesta preocupação com relação à influência da inteligência artificial na tomada decisão das pessoas, questionando se ainda existe controle sobre o livre-arbítrio ante tantos mecanismos de persuasão decorrentes desses novos modelos estatísticos de previsão. O autor exemplifica citando personalizações com objetivo de promover estratégias de comunicação assertivas, como as realizadas pelos algoritmos de inteligência artificial presentes em consultas ao Google, que trazem respostas que variam em função do perfil de quem está buscando a informação. Ou ainda, a rede so-

cial Facebook, que a fim de monetizar seus serviços, declara que seus algoritmos de inteligência artificial são capazes de mapear a personalidade dos usuários com 80% de precisão com base no seu comportamento, abrangendo aspectos como gênero, idade, formação, etnia, orientação sexual, política e religiosa, doenças, uso de substâncias. (Kaufman, 2022). Os exemplos se multiplicam, em se analisando as redes sociais, *marketplaces* e até mesmo buscadores.

Não há como não se refletir, portanto, acerca do alcance, dos riscos e da influência da inteligência artificial na vida da atualidade. A era da hiperconectividade, que fez com que o conceito de meio ambiente fosse ampliado para abranger o meio ambiente digital, apresenta muitos desafios em todos os setores, inclusive na educação e na checagem dos limites de sua utilização.

A utilização das mídias pela infância é corriqueiro e está ocorrendo com precocidade. Pela forma com a qual o desenvolvimento infantil vem sendo compreendido esse crescimento parece ser adversativo. É necessário entender os impactos que as mídias têm na infância, em relação à convivência familiar, inclusive para poder dimensionar quais as possíveis intervenções do direito. Almeida (2021) examinando os impactos da tecnointerferência e as interrupções causadas pela tecnologia para as relações familiares, demonstrou que causam efeitos adversos para as relações conjugais e, na primeira infância, na interação mãe-bebê.

As mídias direcionadas à infância, no Brasil, não são fiscalizadas de modo efetivo, estando, inclusive, imersas em propaganda exacerbada. Os fornecedores de produtos e serviços e o Poder Público não assumem a responsabilidade que deveriam assumir em relação ao excesso de estímulos ao consumo que colocam na Rede.

Nesse sentido, renova-se a importância dos cuidados familiares com o acesso e o conteúdo que os vulneráveis têm ao seu dispor no meio eletrônico. Entretanto, a grande questão que sobressai é de que modo as famílias farão esse controle, ou exercerão atos cuidadosos com as crianças em seus acessos à internet.

Se a vida da criança fora da internet for uma vida de zelo, de presença e de cuidados, parece haver o reconhecimento de que também haverá zelo na vida conectada. Alguns critérios podem ser levados em consideração como, por exemplo, o tempo de exposição à tela, a presença do adulto orientando e controlando os acessos, e a checagem da qualidade do conteúdo visto pelo infante.

A ordem jurídica brasileira designa um dever coletivo de cuidado com a infância. Tomando a questão sob a ótica das mídias digitais e sua interferência aguda na vida da criança e do adolescente, parece que avança o âmbito da mera distração. Por ser um elemento de interferência substancial na vida desses jovens, faz verificar que esse coletivo não é eficaz para o cuidado da infância em nenhuma instância – seja jurídica, social, comportamental, no âmbito psicológico e no de políticas públicas.

Diante disso é que importa ao debate o papel da educação ambiental, na formação da cidadania digital, conceito tratado por Hartmann (2007), associado a outros, como ciber-cidadania, cibersociedade, teledemocracia.

Cidadania digital é a capacidade de usar a tecnologia e as mídias sociais de maneira ética, segura e responsável. A cidadania digital consiste do nível mais básico de maturidade digital e, portanto, pode ser considerado o ponto de partida para uma jornada de transformação da educação. Por outro lado, a educação é o meio mais eficiente para equipar os estudantes com as competências da cidadania digital (Gomes et al, 2021, p. 12).

Latente, pois, a necessidade de efetiva compreensão do atual conceito de meio ambiente, que envolve o meio ambiente digital, e a necessidade de formação de pais e educadores capazes de adequadamente instruir e orientar crianças, jovens e educandos na materialização de uma cidadania digital, em que as pessoas conquistem, além de uma maturidade, também uma ética na sua participação virtual. Somente assim se atingirá uma verdadeira cidadania digital, e a educação ambiental cumprirá seu papel.

Na pesquisa realizada por Gomes et al, abordou-se a educação (Educação 4.0, segundo os autores), tendo sido avaliadas as competências necessárias para o alcance da cidadania digital. Tais habilidades compreenderiam: identidade de cidadão digital íntegra; comportamento online empático, respeitoso e ético; proteção da privacidade digital; e boa reputação digital (Gomes et al, 2021, p. 12). Parecem conceitos auto-explicativos, que envolvem a conduta esperada pelos usuários das redes na prática da cidadania digital, e que não se pretende aprofundar. Apenas considera-se relevante mencionar aspecto igualmente conexo a tudo isso, e que é denominado de letramento digital, mostrando-se de extrema importância para a conquista de uma cidadania verdadeiramente igualitária e inclusiva.

Ainda, defende-se a alfabetização em informação, que visa desenvolver aprendizes ao longo da vida, capacitando-os a encontrar, avaliar e utilizar informações de forma eficaz para resolver problemas e tomar decisões. Uma pessoa alfabetizada em informação é capaz de identificar necessidades de informação, organizá-la, aplicá-la na prática e integrá-la ao conhecimento existente para solucionar problemas (Silva et al, 2005, p. 33).

O uso adequado da informação, aliado ao comportamento on-line que observe empatia, respeito e ética, juntamente com proteção à privacidade e outros valores, demonstram ser fundamentais para a construção da cidadania digital, e para o uso adequado da inteligência artificial. A necessidade de aprimorar-se o uso das tecnologias da informação e comunicação, bem como de construir uma análise crítica das informações obtidas através dessas mesmas tecnologias é que faz emergir a necessidade do letramento digital, conforme defendem Elizete F. V. Conceição e Taís S. Ghisleni (2019), ao citarem a criação, pelo New London Group, de um diagrama que contempla o que entra na pedagogia dos multiletramentos. Seriam requisitos necessários aos sujeitos, para que se considerem digitalmente letrados. Segundo as autoras:

Para tal, o sujeito precisa ser: um “usuário funcional”, abrangendo competência técnica e conhecimento prático, ou seja, saber operar dispositivos e programas e aplicativos; “criador de sentidos”, no qual o sujeito precisa ser mais produtor do que consumidor, mas para ser criador de sentidos, ele precisa ser “analista crítico”, precisa ler criticamente as informações disponíveis no digital, somente para então criar sentidos, e que tais se transformem, usando o que foi aprendido de novos modos, ou seja, utilizando-se da ideia do remix, criar a partir do que os outros já produziram, assim tornando-se um sujeito “transformador” (Conceição and Ghisleni, 2019, p. 13).

Para a concretização da cidadania digital, é necessário o desenvolvimento de um letramento digital, em que os cidadãos, além de disporem de capacidade de operação dos dispositivos tecnológicos e respectivos programas e aplicativos, também sejam capazes de tornar-se agentes transformadores da sociedade. Para isso, devem os cidadãos ter capacidade de analisar criticamente a informação recebida, produzindo novas ideias, criando e disseminando novos sentidos. Assim se conquistará uma evolução cada vez maior da sociedade, e se minimizarão os riscos da inteligência artificial, reconhecendo-se a efetividade da educação ambiental.

5. Considerações finais

Certo é que a concepção de meio ambiente, por sua amplitude, envolve o meio digital, caracterizado e resultado da hiperconectividade a que a sociedade contemporânea está inserida. Não se cogita a vida sem a tecnologia ou desprovida de recursos e sistemas inteligentes.

Desde softwares assistentes pessoais, presentes em smartphones, computadores, e outros dispositivos, até utensílios domésticos, aplicativos para crianças e complexos sistemas de gestão empresarial e de serviços públicos, vê-se a inteligência artificial operando em prol de uma efetividade e análise assertiva de dados. Mas não são apenas benefícios que se verificam, mas inúmeros riscos, gerados por discriminação algorítmica, inadequada influenciação de ideias e

conteúdos, e outros tantos que sequer foi possível mensurar na pesquisa.

O exercício da cidadania, hoje, tornou-se ainda mais facilitado e acessível em razão dos mecanismos tecnológicos e da simplificação no uso de meios de comunicação presentes na Internet. Todos operam por inteligência artificial, e possibilitam a livre manifestação do pensamento e a busca da informação. Entretanto, algoritmos direcionam conteúdo de acordo com as preferências do usuário, o que, além de poder provocar a formação de verdadeiras bolhas informacionais, também incentivam o excesso de exposição, especialmente de crianças e adolescentes, às telas.

Inegável é que se passa cada vez mais tempo conectado a dispositivos tecnológicos, o que parece ser ainda mais grave no caso das gerações que já nasceram inseridas no universo tecnológico.

Esta realidade desafia o Direito, mas acima de tudo, desafia as famílias e os educadores, que, para além de formar cidadãos, precisam preparar verdadeiros cidadãos digitais, preparados não para exilar-se das tecnologias, mas aprender a com elas conviver, tornando-as ferramentas de sua evolução. Somente aliado à tecnologia é que o homem minimizará o risco de ser por ela substituído.

Por fim, este parece ser o papel da educação ambiental: orientar os educandos sobre os mecanismos tecnológicos e digitais disponíveis, auxiliando-os na adoção de análise crítica e responsável, para que não apenas operem ou disseminem informações, mas sejam capazes de gerar novos sentidos e compartilhá-los. Capacidade para isso, somente o homem possui, razão pela qual, amparado na educação, jamais se deixará substituir pela máquina.

Referências

- Almeida, M. L. (2021) *O uso de mídias digitais na primeira infância: Tecnointerferência, variáveis associadas ao uso e proposta de intervenção*, Tese (Doutorado em Psicologia), Universidade Federal do Rio Grande do Sul, Porto Alegre.
- Cavedon, R, Ferreira, H. S. and Freitas, C. O. de A. (2015) 'O meio ambiente digital sob a ótica da Teoria da Sociedade de Risco: os avanços da informática em debate', *Revista Direito Ambiental e sociedade*, 5(1), pp. 194-223.
- Conceição, E. F. V. and Ghisleni, T. S. (2019), 'Era digital: letramento(s) digital(is)', *Revista Research, Society And Development*, 8(12), e398121785 [online]. Available at: [https://rsdjournal.org/index.php/rsd/article /view/1785](https://rsdjournal.org/index.php/rsd/article/view/1785) (Accessed: 10 August 2023).
- Doneda, D. C. M. et al (2018) 'Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal', *Pensar: Revista de Ciências Jurídicas*, Fortaleza, 23(4), p. 1-17, October/December.
- Fiorillo, C. A. P. and Souza, C. F. (2016) 'O Direito à Comunicação e a Tutela do Meio Ambiente Digital', *Revista de Direito Ambiental e Socioambientalismo*, 2, July/December.
- Fisher, M. (2023) *A Máquina do Caos: Como as Redes Sociais Reprogramaram Nossa Mente e Nosso Mundo*. São Paulo: Todavia.
- Frazão, A. Discriminação algorítmica: por que os algoritmos preocupam quando acertam e quando erram? *Ana Frazão* [online]. Available at: http://www.professoraanafrazao.com.br/files/publicacoes/2021-08-04-1Discriminacao_algoritmica_por_que_os_algoritmos_preocupam_quando_acertam_e_quando_erram_Mapeando_algunas_das_principais_discriminacoes_algoritmicas_ja_identificadas_Parte_VIII.pdf (Accessed: 10 August 2023).
- Freitas, J. and Freitas, T. B. (2020) *Direito e Inteligência Artificial: em defesa do humano*. Belo Horizonte: Fórum.

- Gomes, G. F., Oliveira, K. S. and Souza, R. A. C. (2021) 'Competências da Cidadania Digital: especificação e avaliação de uma proposta de experiência de ensino-aprendizagem', *Revista Gestão.org*, 19 [online]. Available at: <https://periodicos.ufpe.br/revistas/gestaoorg/article/view/252647/40150> (Accessed: 10 August 2023).
- Guattari, F. (1990) *Três ecologias*. Translated by Maria Cristina F. Bittencout. Campinas: Papirus.
- Harari, Y. (2018) *21 lições para o século 21*. São Paulo: Companhia das Letras.
- Hartmann, I. A. M. (2007) 'O Acesso à Internet como Direito Fundamental', *Portal e-gov* [online]. Available at: https://egov.ufsc.br/portal/sites/default/files/ivar_hartmann.pdf (Accessed: 10 August 2023).
- Kaufman, D. (2022) *Desmistificando a Inteligência Artificial*. Belo Horizonte: Autêntica.
- Lacerda, B. T. Z. (2022) *Estatuto jurídico da Inteligência Artificial: entre categorias e conceitos, a busca por marcos regulatórios*. Indaiatuba: Foco.
- Magrani, E. (2018) *A internet das coisas*. Rio de Janeiro: FGV Editora.
- Markus, G. (2021) *O sentido do pensar: a filosofia desafia a inteligência artificial*. Translated by Lucas Machado. Petrópolis: Vozes.
- Monk, S. (2018) *Internet das Coisas: Uma Introdução com o Photon* – Série Tekne. Translated by Anatólio Laschuk. Porto Alegre: Bookman.
- Onody, R. (2021) 'Teste de Turing e Inteligência Artificial', *Universidade de São Paulo, Instituto de Física de São Paulo* [online]. Available at: <https://www2.ifsc.usp.br/portal-ifsc/teste-de-turing-e-inteligencia-artificial/> (Accessed: 10 August 2023).
- Rech, A. U. (2020) *Inteligência artificial, meio ambiente e cidades inteligentes*. Caxias do Sul: EducS.

- Silva, H., Jambreiro, O., Lima, J. and Brandão, M. A. (2005) 'Inclusão digital e educação para a competência informacional: uma questão de ética e cidadania', *Ciência da Informação*, 34(1), January/April.
- Turing, A. (1950) 'Computing machinery and intelligence', Oxford, Oxford University Press, *Mind, New Series*, 59(236), pp. 433-460 [online]. Available at: <https://phil415.pbworks.com/f/TuringComputing.pdf> (Accessed: 10 August 2023).

SOBRE OS AUTORES

AFONSO VINÍCIO KIRSCHNER FRÖHLICH: Mestre em Direito Público pela Universidade do Vale do Rio dos Sinos (Unisinos). Bacharel em Direito pela mesma Universidade. Especialista em Direito e Negócios Imobiliários pela Fundação Escola Superior do Ministério Público (FMP). Advogado sócio do Escritório de Advocacia Afonso Fröhlich Advogados Associados. E-mail: afonsovinicio@afrohlich.adv.br e afonsovkf@gmail.com.

ALEXANDRE CORTEZ FERNANDES: Doutor em educação, professor da graduação e do PPG em Direito da Universidade de Caxias do Sul (UCS). E-mail: acfernann@ucs.br.

CARLOS DE CORES HELGUERA: Doctor em Derecho por la Universidad Pontificia Comillas, Madrid. Director de Derecho Privado de la Universidad Católica del Uruguay.

CRISTIANO COLOMBO: Pós-Doutor em Direito, Pontificia Universidade Católica do Rio Grande do Sul (PUCRS). Doutor e Mestre em Direito, Programa de Pós-Graduação em Direito da Universidade Federal do Rio Grande do Sul (UFRGS). Professor Permanente do Mestrado Profissional em Direito da Empresa e dos Negócios da Unisinos; Professor de Graduação em Direito e Relações Internacionais da Unisinos; Professor de Graduação em Direito da Faculdade Verbo Jurídico; Pesquisador Fapergs (2021-2023). E-mail: cristianocolombo@unisinos.br.

DANY STEVEN GÓMEZ AGUDELO: Magíster en Derecho de la Universidad Católica Luis Amigó. Docente de Derecho Informático y Nuevas Tecnologías, Derecho Procesal Administrativo, coordinador del área de Derecho Público del Programa de Derecho y líder del grupo de investigación *Jurídicas y Sociales*, categoría A en Minciencias, en la Universidad Católica Luis Amigó, Medellín, Colombia. Correo electrónico: dany.gomezag@amigo.edu.co y danygomezagudelo@hotmail.com. ORCID: <https://orcid.org/0000-0003-2687-0146>.

DARCI GUIMARÃES RIBEIRO: Advogado e Árbitro. Pós-Doutor em Direito Processual Constitucional pela *Università degli Studi di Firenze*. Doutor em Direito pela *Universitat de Barcelona*. Especialista e Mestre pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Professor Titular de Direito Processual Civil da Unisinos e PUCRS. Professor Adjunto de Direito Processual Civil da UFN. Professor do Programa de Pós-Graduação em Direito da Unisinos (Mestrado, Doutorado e Pós-Doutorado). Membro da *International Association of Procedural Law*. Membro do Instituto Ibero-americano de Derecho Procesal. Membro do Instituto Brasileiro de Direito Processual Civil. Membro da Academia Brasileira de Direito Processual. Membro da Asociación Argentina de Derecho Procesal. Huésped de Honor da Casa de Altos Estudios da Universidad Nacional de Córdoba, Argentina. Membro representante do Brasil no Projeto Internacional de Investigação sobre *Nuevos retos tecnológicos del derecho probatorio*, (PID2020-115304GB-C21) financiado pelo Ministério de Ciência e Inovação – MEC – da Espanha. E-mail: darci.guimaraes@terra.com.br.

DIOGO DE CALASANS MELO ANDRADE: Doutor em direito político e econômico pela Universidade Mackenzie. Mestre em direito, na área de constitucionalização do direito pela UFS. Professor titular da graduação e do mestrado e doutorado em direitos humanos do PPGD-UNIT. Líder do grupo de pesquisa “Novas tecnologias e o impacto nos Direitos Humanos” do mestrado em direito Humanos da UNIT, advogado. E-mail: contato@profdiogocalasans.org.

GUILHERME DAMASIO GOULART: Doutor e Mestre em Direito pela Universidade Federal do Rio Grande do Sul (UFRGS). Atua como advogado e consultor em Segurança da Informação, Direito da Tecnologia e Proteção de Dados Pessoais. Professor de Direito Civil do Cesuca. E-mail: guilherme@direitodatecnologia.com.

GUSTAVO DA SILVA SANTANNA: Doutor e Mestre em Direito pela Unisinos. Professor da Universidade Federal do Rio Grande do Sul – UFRGS – e da Atitus Educação. Professor das Especializações em Direito do Estado da Universidade Federal do Rio Grande do Sul – UFRGS, em Direito Digital, Gestão da Inovação e Propriedade Intelectual e Direito Administrativo da Pontifícia Universidade Católica de Minas Gerais – PUC/MINAS. Procurador do Município de Alvorada/RS. Advogado. ORCID: <https://orcid.org/0000-0003-1190-3355>. Lattes: <http://lattes.cnpq.br/3394849903197236>. E-mail: gssantanna@hotmail.com.

HAIDE MARIA HUPFFER: Pós-Doutora em Direito pela Unisinos. Doutora e Mestre em Direito pela Unisinos; Docente e Pesquisadora no Programa de Pós-Graduação em Qualidade Ambiental e no curso de Direito da Universidade Feevale. Líder do Grupo de Pesquisa Direito e Desenvolvimento CNPq/FEEVALE. Coordenadora do Projeto “Inteligência Artificial no Sul Global: Regulação, Riscos Discriminatórios, Governança e Responsabilidades” financiado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (Fapergs). ORCID: <https://orcid.org/0000-0002-4965-9258>. Lattes: <http://lattes.cnpq.br/4950629941533824>. E-mail: haide@feevale.br.

JOSÉ JORGE BARRETO TORRES: Doutorando em Direitos Humanos pela Universidade Tiradentes. Mestre em Ciência da Computação pela Universidade Federal de Sergipe e Especialista em Bancos de Dados pela Universidade Tiradentes. Bacharel em Direito e Sistemas de Informação pela Universidade Tiradentes. E-mail: doutorado_jorgebt@souunit.com.br.

JOSÉ LUIZ DE MOURA FALEIROS JÚNIOR: Doutorando em Direito Civil pela Universidade de São Paulo – USP/Largo de São Francisco. Doutorando em Direito, na área de estudo “Direito, Tecnologia e Inovação”, pela Universidade Federal de Minas Gerais – UFMG. Mestre e Bacharel em Direito pela Universidade Federal de Uberlândia – UFU. Especialista em Direito Digital. Especialista em Direito Civil e Empresarial. Advogado e Professor. E-mail: jfaleiros@usp.br.

JOSÉ LUIZ DE MOURA FALEIROS JÚNIOR: Doutorando em Direito Civil pela Universidade de São Paulo – USP/Largo de São Francisco. Doutorando em Direito, na área de estudo ‘Direito, Tecnologia e Inovação’, pela Universidade Federal de Minas Gerais – UFMG. Mestre e Bacharel em Direito pela Universidade Federal de Uberlândia – UFU. Especialista em Direito Digital. Especialista em Direito Civil e Empresarial. Advogado e Professor. ORCID: <https://orcid.org/0000-0002-0192-2336>. E-mail: jfaleiros@usp.br.

MARTÍN JULIÁN ACEVEDO – MIÑO: Abogado (UCSF) y doctor en Ciencias Jurídicas (UCA). Profesor de Derecho Constitucional (UCA). Profesor de doctorado (UCSF). Diputado constituyente estadual (Entre Ríos, 2008). Conjuez del Superior Tribunal de Justicia de Entre Ríos (2020/2022). Secretario General de la Asociación Argentina de Derecho Constitucional. ORCID: <https://orcid.org/0000-0001-9855-7576>.

NELSON ARTEAGA BOTELLO: Doctor en Sociología. Profesor investigador de la Facultad Latinoamericana de Ciencias Sociales, sede México. Miembro del Sistema Nacional de Investigadores Nivel III. Miembro de la Academia Mexicana de Ciencias. Entre sus últimas publicaciones se encuentran: *Semantics of Violence: Revolt and Political Assassination in Mexico*, Palgrave-MacMillan. *Sociología y Ciencia Ficción: Philip K. Dick más allá de Orwell y Foucault*. Flacso-México (2020). Correo electrónico: nelson.arteaga@flacso.edu.mx.

PAOLA CANTARINI-GUERRA: Doutorado em filosofia do direito (Universidade do Salento, Itália); mestrado e doutorado em direito e em filosofia (PUC-SP); pós-doutorado em Filosofia, pensamento crítico e arte (European Graduate School), em Direito FD-USP), em sociologia jurídica (Universidade de Coimbra), em Tecnologias da Inteligência e do Design Digital e em Inteligência Artificial e Filosofia (Instituto de Estudos Avançados – USP – Cátedra Oscar Sala). Fundadora e presidente do Instituto EthikAI. E-mail: paolacantarini@gmail.com.

PATRÍCIA MONTEMEZZO: Mestre em direito, professora de Direito da Universidade de Caxias do Sul (UCS) e advogada. E-mail: pmonteme@ucs.br.

PATRÍCIA VERÔNICA NUNES CARVALHO SOBRAL DE SOUZA: Pós-Doutora em Direito pela Mediterranea International Centre for Human Rights Research della Università Mediterranea di Reggio Calabria (Itália). Pós-Doutora em Direito e Doutora em Direito Público pela Universidade Federal da Bahia-UFBA. Professora Titular de Graduação e do Programa de Pós-graduação da Universidade Tiradentes. E-mail: patncss@gmail.com.

SEBASTIÁN DÍAZ BOLÍVAR: Magíster en Derecho Procesal Contemporáneo. Doctorando en Derecho Procesal Contemporáneo de la Universidad de Medellín. Coordinador del área de derecho privado civil II del Programa de Derecho de la Universidad Católica Luis Amigó. Miembro del grupo de investigación *Jurídicas y Sociales*, categoría A en Minciencias, en la Universidad Católica Luis Amigó, Medellín, Colombia. Correo electrónico: sebastian.diazbo@amigo.edu.co. ORCID: <https://orcid.org/0000-0001-7314-6412>.

TAÍS FERNANDA BLAUTH: Graduada em Direito pela Universidade Feevale e é mestre em política e relações internacionais pela Durham University (Reino Unido). Atualmente é doutoranda em Inteligência Artificial e Relações Internacionais na University of Groningen/Campus Fryslân (Países Baixos). Além disso, é pesquisadora do Data Research Center, Humboldt Institute for Internet and Society (Alemanha), e Information Society Law Center (Itália). Seus interesses de pesquisa incluem governança de IA, inovação militar e controle de armas.

WILLIS SANTIAGO GUERRA FILHO: Professor Titular do Centro de Ciências Jurídicas e Políticas da Universidade Federal do Estado do Rio de Janeiro (UNIRIO). Livre-docente em Filosofia do Direito pela Universidade Federal do Ceará (UFC); Doutor em Ciência do Direito pela Universidade de Bielefeld, Alemanha; Doutor e Pós-Doutor em Filosofia pela Universidade Federal do Rio de Janeiro (UFRJ); Mestre em Direito, Doutor em Comunicação e Semiótica e em Psicologia Social/Política (PUC-SP). E-mail: willis.filho@unirio.br.

WILSON ENGELMANN: Doutor e Mestre em Direito Público, Programa de Pós-Graduação em Direito da Universidade do Vale do Rio dos Sinos – Unisinos, Brasil; realizou Estágio de Pós-Doutorado em Direito Público – Direitos Humanos, no Centro de Estudios de Seguridad (CESEG) da Universidade de Santiago de Compostela, Espanha; Professor e Pesquisador do Programa de Pós-Graduação em Direito – Mestrado e Doutorado e do Mestrado Profissional em Direito da Empresa e dos Negócios, ambos da Unisinos; Bolsista de Produtividade em Pesquisa do CNPq; Líder do Grupo de Pesquisa JusNano. ORCID: <https://orcid.org/0000-0002-0012-3559>. Lattes: <http://lattes.cnpq.br/7143561813892945>. E-mail: wengelmann@unisinos.br.

ÍNDICE REMISSIVO

A

Acción de tutela 7, 239, 240, 241, 248, 250, 251, 254
Accountability 84, 99, 151, 157, 166, 167
América Latina 111, 113, 133, 260, 262, 263, 270
Aprendizado de máquina 18, 47, 80, 215, 283, 284, 289
Aprendizado profundo 80, 108
Autoimunidade 111, 112, 128, 129, 130
Autopoiese 129, 130, 134
Autorregulação regulada 14, 24, 26, 29, 31

B

Big Data 47, 48, 50, 52, 59, 70, 71, 78, 80, 83, 91, 102, 191,
240, 291

C

Capitalismo de vigilância 45, 47, 48, 49, 62, 66, 68
Ciber-bulling 173, 177
Ciberescravidão 6, 111, 112, 127, 130
Ciberespaço 52, 275
Cidades Digitais 260, 261, 263, 264, 266, 267, 268, 273, 274,
275, 276, 277
Cidades Inteligentes 7, 259, 260, 261, 262, 263, 264, 268, 269,
270, 271, 273, 274, 276, 277, 278, 298
Cidades Sustentáveis 261, 262, 264, 265, 266, 274, 275, 276
Colonialidade 52, 70, 71
Colonialismo digital 6, 11, 45, 46, 47, 49, 54, 55, 62, 66
Crimen organizado 177, 178

D

Dados pessoais 27, 52, 54, 63, 65, 88, 108, 137, 140, 154, 290
Darwin 221, 230, 231, 232, 237, 288, 289
Deep learning 39, 47, 57, 80, 83, 84, 107, 108, 284, 288
Dever de fundamentar 7, 193, 212

Dignidad de las personas 185
Dignidade da pessoa 66
Direito autoral 146, 164, 165
Direito de não ser discriminado 93
Direitos do autor 160
Direitos fundamentais 12, 14, 20, 23, 25, 26, 60, 61, 69, 71,
72, 76, 77, 96, 97, 98, 100, 113, 133
Direitos humanos 11, 13, 14, 20, 23, 24, 25, 29, 46, 47, 52, 55,
59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 73, 75, 77, 82,
84, 91, 92, 94, 95, 97, 98, 101, 113, 133, 259, 275, 302,
303, 306, 311
Discriminação algorítmica 6, 45, 46, 297

E

Equidade e Justiça 12, 27

F

Fascinación bigdataísta 39
Finnis 14, 15, 21, 22, 29, 31
Fundamentação das decisões 194, 195, 196, 205, 206, 207,
212, 213

G

Global lex digitalis 7, 221, 224, 225, 226, 228, 229, 235, 236,
237
GPDR 93

H

Harari 63, 69, 221, 225, 226, 227, 228, 229, 230, 231, 232,
234, 238, 289, 298
Heidegger 29, 116, 119, 124, 125, 126, 127, 128, 133, 134
Hibridismos 115
Hiperconectividade 25, 50, 279, 280, 281, 282, 283, 292, 295
Homo Deus 221, 222, 225, 228, 229, 230, 236, 238
Homo sapiens 154, 221, 225, 228, 235, 236, 237

I

IA Generativa 81, 193, 194, 196, 198, 200, 202, 203, 204, 206, 207, 208, 209, 210, 211, 212, 213
 IA responsável 12, 14, 15, 18, 22, 28
 Imunologia social 129, 130
 Indígenas 181, 183
 Inforgs 45, 46, 47
 Infosfera 25, 45, 47, 49, 54
 Inteligência Artificial Generativa 7, 193, 195, 197, 200, 205, 207, 212

J

Justiça algorítmica 84, 90, 92, 102, 212

L

Large Language Models 6, 31, 137, 138, 139, 140, 157, 1661, 167
 LGPD 63, 65
 LLM 137, 140, 141, 143, 144, 147, 149, 150, 151, 156, 157, 158
 Luhmann 128, 129, 131, 132, 135

M

Machine Learning 39, 47, 51, 69, 75, 78, 79, 80, 84, 91, 92, 102, 107, 108, 142, 198, 199, 243, 284
 Minorias 83

N

Neurodireitos 6, 11, 111, 112, 113, 114, 115, 120, 123, 126, 130, 132
 Neurotecnologia 81, 113
 Norte Global 45, 47, 54, 66, 112
 Novo Constitucionalismo Latino-americano 111, 112

P

Pensamento descolonial 46, 49, 54, 55, 73
 Perfilização 46, 48
 Phrónesis aristotélica 14, 15, 18
 Pluralismo jurídico 47, 51, 52, 53, 60, 73

Princípios 12, 13, 14, 15, 22, 23, 24, 25, 26, 27, 28, 29, 31, 59,
92, 93, 94, 97, 98, 99, 100, 101, 110, 118, 132, 135, 139,
156, 283
Privacidade 11, 12, 27, 54, 63, 81, 83, 89, 91, 98, 100, 101,
102, 103, 104, 109, 294
Protección de las víctimas 183

S

Segurança Pública 7, 102, 259, 260, 262, 263, 270, 271, 272,
273, 274, 275, 276, 277, 287
Sociología cultural 171, 186
Sul Global 11, 111, 112, 303, 311
Supervisão humana 12, 27, 98
Supremacia constitucional 37, 38
Sustentabilidade 95, 98, 100, 104, 259, 260, 261, 264, 265,
268, 269, 270

T

Tanatopolítica 130
Techné 128
Teoría de la evolución 7, 221, 231, 232
Transparência 12, 27, 76, 77, 83, 89, 91, 92, 94, 98, 99, 100,
101, 102, 103, 113, 146, 151, 152, 154, 206

V

Viés discriminatório 84
Vieses discriminatórios 75, 76, 77, 85, 103
Violencia contra las mujeres 177, 179, 180
Violencias en las redes sociales 6, 171



CASA LEIRIA
Rua do Parque, 470
São Leopoldo-RS Brasil
casaleiria@casaleiria.com.br

ORGANIZADORES:

HAIDE MARIA HUPFFER: Pós-Doutora em Direito pela Unisinos. Doutora e Mestre em Direito pela Unisinos; Docente e Pesquisadora no Programa de Pós-Graduação em Qualidade Ambiental e no curso de Direito da Universidade Feevale. Líder do Grupo de Pesquisa Direito e Desenvolvimento CNPq/Feevale. Coordenadora do Projeto “Inteligência Artificial no Sul Global: Regulação, Riscos Discriminatórios, Governança e Responsabilidades” financiado pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (Fapergs).

WILSON ENGELMANN: Doutor e Mestre em Direito Público, Programa de Pós-Graduação em Direito da Universidade do Vale do Rio dos Sinos - Unisinos, Brasil; realizou Estágio de Pós-Doutorado em Direito Público – Direitos Humanos, no Centro de Estudios de Seguridad (CESEG) da Universidade de Santiago de Compostela, Espanha. Professor e Pesquisador do Programa de Pós-Graduação em Direito – Mestrado e Doutorado e do Mestrado Profissional em Direito da Empresa e dos Negócios, ambos da Unisinos. Bolsista de Produtividade em Pesquisa do CNPq. Líder do Grupo de Pesquisa JusNano.

TAÍS FERNANDA BLAUTH: Graduada em Direito pela Universidade Feevale e é mestre em política e relações internacionais pela Durham University (Reino Unido). Atualmente é doutoranda em Inteligência Artificial e Relações Internacionais na University of Groningen/Campus Fryslân (Países Baixos). Além disso, é pesquisadora do Data Research Center, Humboldt Institute for Internet and Society (Alemanha), e Information Society Law Center (Itália). Seus interesses de pesquisa incluem governança de IA, inovação militar e controle de armas.

A promessa da IA é que ela melhore nossa vida, pois esses algoritmos podem, teoricamente, decidir sem vieses de gênero, raça ou classe, de forma lógica e racional, evitando qualquer parcialidade e discriminação. Entretanto, pode ser que isso não ocorra tal como esperado, uma vez que todas essas decisões referidas anteriormente, inevitavelmente, parecem envolver padrões éticos e avaliações morais complexas, o que pode resultar em injustiças. Por exemplo, os carros autônomos devem sempre se esforçar para minimizar as baixas, mesmo que às vezes isso signifique sacrificar seus próprios passageiros para um bem maior? As armas autônomas letais devem sempre objetivar a vitória, mesmo ao custo de eliminar um alvo civil ou um soldado ferido, o que colocaria em risco a dignidade humana? As crianças devem sempre ter prioridade para transplantes de órgãos, mesmo quando um paciente mais velho é uma combinação genética melhor para um órgão disponível? Os algoritmos usados em tribunais devem sempre procurar reduzir a reincidência, mesmo resultando em uma discriminação injusta para os réus negros?

Denis Coitinho

Coordenador Executivo e Professor do
Programa de Pós-Graduação em Filosofia da Unisinos

