Inteligência artificial: promessas, riscos e regulação. Algo de novo debaixo do sol[2]

Artificial intelligence: promises, risks and regulation. Something New Under the Sun

Inteligencia artificial: promesas, riesgos y regulación. ¿Algo nuevo bajo el sol?

Luís Roberto Barroso

Presidente do Supremo Tribunal Federal. Professor titular da Universidade do Estado do Rio de Janeiro (Uerj) e do Centro Universitário de Brasilia (Uniceub). Doutor e livre docente pela Uerj. Mestre pela Yale Law School, EUA. *Visiting scholar* na Faculdade de Direito de Harvard. *Senior fellow* da Harvard Kennedy School, EUA.

Patrícia Perrone Campos Mello

Secretária de Altos Estudos do Supremo Tribunal Federal. Professora de direito constitucional da Uerj e do Uniceub. Doutora e mestre em direito pela Uerj. Estudos de pós-doutorado como *visiting scholar* no Instituto Max Planck de Direito Público Comparado e Internacional, Alemanha, e na Harvard Kennedy School, EUA. Procuradora do Estado do Rio de Janeiro.

2 Artigo publicado na Revista *Direito e Práxis*, v. 15, n. 4, 2024.

SUMÁRIO: Nota inicial; Introdução: O alvorecer da quarta revolução industrial; 1. Um admirável mundo novo; 2. O que é inteligência artificial; 3. Aprendizado de máquina, modelo fundacional e outros conceitos relevantes; Parte I: A inteligência artificial e seus benefícios; 1. Melhor capacidade decisória em muitas áreas; 2. Automação; 3. Linguagem; 4. Pesquisa e inovação; 5. Aplicações na medicina; 6. Aplicações no sistema de justiça; 7. Educação e cultura; 8. Outras aplicações úteis da IA; 8.1 Utilidades práticas do dia a dia; 8.2 Proteção ao meio ambiente; 8.3 Personalização das relações comerciais e outras; Parte II: A inteligência artificial e seus riscos; 1. Impacto sobre o mercado de trabalho; 2. Utilização para fins bélicos; 3. Massificação da desinformação; 4. Violação da privacidade; 5. Discriminação algorítmica; 6. Questões sobre propriedade intelectual e direitos autorais; Parte III: Alguns princípios para regulação da inteligência artificial; 1. Complexidades da regulação; 2. Alguns esforços de regulação; 3. Algumas diretrizes; 3.1 Defesa dos direitos fundamentais; 3.2 Proteção da democracia; 3.3 Promoção da boa governança; Conclusão; Referências.

RESUMO: O presente artigo trata das potencialidades e riscos da inteligência artificial (IA). Com esse objetivo, posiciona a IA no alvorecer da IV Revolução Industrial, explica suas categorias essenciais e modo de operar. Aborda os benefícios trazidos pela nova tecnologia: ampliação da capacidade decisória humana, automação, avanços em pesquisa e inovação, medicina e educação, entre outros. Examina os riscos que ela gera, entre os quais: impactos sobre o mercado de trabalho, utilização para fins bélicos, massificação da desinformação e violação a direitos fundamentais. Propõe princípios para a regulação da IA. Demonstra que se trata de uma tecnologia com grande potencial, cujos efeitos reais dependerão sobretudo do uso que faremos dela. Em tais condições, o direito tem o importante desafio de produzir um desenho institucional que incentive seu bom uso e contenha o seu desvirtuamento.

PALAVRAS-CHAVE: inteligência artificial; direitos fundamentais; democracia; trabalho; riscos; regulação.

TABLE OF CONTENTS: Initial Note; Introduction: The dawn of the fourth industrial revolution; 1. A brave new world; 2. What artificial intelligence is; 3. Machine learning, foundation models, and other relevant concepts; Part I: Artificial intelligence and its benefits; 1. Better decision-making in many áreas; 2. Automation; 3. Language; 4. Research and innovation; 5. Applications in medicine; 6. Applications in the justice system; 7. Education and culture; 8. Other useful applications of AI; 8.1 Practical everyday utilities; 8.2 Environmental protection; 8.3 Personalization of commercial and other relationships; Part II: Artificial intelligence and its risks; 1. Impact on the job Market; 2. Use for military purposes; 3. Mass dissemination of disinformation; 4. Violation of privacy; 5. Algorithmic discrimination; 6. Issues regarding in-

tellectual property and copyright; Part III: Some principles for regulating artificial intelligence; 1. Complexities of regulation; 2. Some regulatory efforts: 3. Some guidelines; 3.1 Protection of fundamental rights; 3.2 Protection of democracy; 3.3 Promotion of good governance; Conclusion; References
ABSTRACT: This article deals with the benefits and risks of artificial intelligence (AI). It places AI at the dawn of the IV Industrial Revolution, explains its essential categories and way of operating. It addresses the benefits brought by this new technology: expansion of human decision-making capacity, automation, research and innovation, medical school and education, among others. It examines the risks it presents, concerning impacts on the labor market, use for military purposes, dissemination of misinformation and violation of fundamental rights. Finally, it proposes principles for the regulation of AI. The paper demonstrates that AI is a technology with great potential, whose real effects depend on the use we make of it. In such conditions, the law and its operators must look for an institutional design that encourages its positive use and contains its distortion.
KEYWORDS: artificial intelligence; fundamental rights; democracy; labor market; risks; regulation.

CONTENIDO: Nota inicial; Introducción: El amanecer de la cuarta revolución industrial; 1. Un mundo feliz y nuevo; 2. Qué es la inteligencia artificial; 3. Aprendizaje automático, modelos fundacionales y otros conceptos relevantes; Parte I: La inteligencia artificial y sus benefícios; 1. Mejor capacidad de decisión en muchas áreas; 2. Automatización; 3. Lenguaje; 4. Investigación e innovación; 5. Aplicaciones en la medicina; 6. Aplicaciones en el sistema de justicia; 7. Educación y cultura; 8. Otras aplicaciones útiles de la IA; 8.1 Utilidades prácticas del día a día; 8.2 Protección del medio ambiente; 8.3 Personalización de las relaciones comerciales y otras; Parte II: La inteligencia artificial y sus riesgos; 1. Impacto sobre el mercado laboral; 2. Uso con fines bélicos; 3. Masificación de la desinformación; 4. Violación de la privacidad; 5. Discriminación algorítmica; 6. Cuestiones sobre propiedad intelectual y derechos de autor; Parte III: Algunos principios para la regulación de la inteligencia artificial; 1. Complejidades de la regulación; 2. Algunos esfuerzos de regulación; 3. Algunas directrices; 3.1 Defensa de los derechos fundamentales; 3.2 Protección de la democracia; 3.3 Promoción de la buena gobernanza; Conclusión; Referencias.
RESUMEN: El presente artículo trata sobre las potencialidades y riesgos de la inteligencia artificial (IA). Con ese objetivo, sitúa a la IA en el amanecer de la IV Revolución Industrial, explicando sus categorías esenciales y su modo de funcionamiento. Aborda los beneficios que aporta esta nueva tecnología: ampliación de la capacidad de decisión humana, automatización, avances en investigación e innovación, medicina y educación, entre otros. Examina los

riesgos que genera, entre ellos: impactos sobre el mercado laboral, uso con fines bélicos, masificación de la desinformación y violación de derechos fundamentales. Propone principios para la regulación de la IA. Demuestra que se trata de una tecnología con gran potencial, cuyos efectos reales dependerán, sobre todo, del uso que hagamos de ella. En tales condiciones, el derecho tiene el importante desafío de crear un diseño institucional que fomente su buen uso y contenga su desviación.
PALABRAS CLAVE: inteligencia artificial; derechos fundamentales; democracia; trabajo; riesgos; regulación.

Nota inicial[3]

Há algo de novo debaixo do sol[4]. Muitas de nossas crenças e certezas podem estar com os dias contados. Como os antigos navegadores contemplavam a imensidão dos oceanos, repletos de promessas, mistérios e perigos, estamos novamente de frente para um mundo desconhecido. Paira no ar a sensação de que uma transformação profunda está por vir. Uma revolução, talvez. Algo grandioso como a invenção da prensa por tipos móveis, que difundiu exponencialmente o conhecimento humano, ou o Iluminismo, que reformulou a vida social, a cultura e a política (Kissinger; Schmidt; Huttenlocher, 2023). O futuro nunca pareceu tão próximo e imprevisível[5].

Diante das possibilidades aparentemente infinitas da tecnologia, só existe uma carta de navegação segura: os valores que desde muito longe devem pautar o avanço civilizatório e a evolução da condição humana na Terra. Laicos ou míticos, eles vêm da Grécia, passam pela Torá, pelos Evangelhos, Buda, Tomás de Aquino, Kant e muitos outros que construíram o patrimônio ético da humanidade. Mas há um ponto dramático aqui: o vertiginoso progresso científico que assistimos, cumulativamente, ao longo dos séculos, não

3 Os autores são gratos a Pedro Henrique Ribeiro Morais e Silva pela ajuda na pesquisa e por sugestões na elaboração do texto, e, também, a Frederico Alvim, por importantes comentários e recomendações bibliográficas.

4 Há uma passagem bíblica conhecida, contida em *Eclesiastes* 1:9, na qual se lê: "O que foi é o que há de ser; o que se fez, isso se tornará a fazer. Não há nada de novo debaixo do sol". O significado dessa frase é que passam as eras e os seres humanos se debatem com as mesmas questões existenciais. Talvez algumas novas questões estejam surgindo, no entanto.

5 Essa imprevisibilidade é explicável, em parte, pela progressiva autonomia adquirida pelas soluções baseadas em aprendizado de máquina e, em parte, pelo caráter acelerado das inovações nesse campo, que se sucedem (ou se renovam) em uma marcha que prejudica a sua plena compreensão. Nina Schick (2020, p. 11) observa que houve quatro séculos entre a invenção da prensa e o desenvolvimento da fotografia, por exemplo, mas, em apenas três décadas, passamos da eclosão da internet aos *smartphones*, e daí para a plataformização das vidas nas redes sociais, com sérias implicações no regime de informações. Segundo a autora, mudanças tão rápidas em segmentos vitais acarretam um alto componente de incerteza, que deve ser ponderado pela sociedade como um todo.

tem sido acompanhado de uma correspondente evolução ética – e mesmo espiritual – da condição humana. O bem, a justiça real e a solidariedade são frequentemente negligenciados num mundo de pobreza extrema em muitos lugares, desigualdades injustas, guerras e uma ordem doméstica e internacional em que alguns ganham todas e outros perdem sempre. É nesse cenário que se coloca o tema da inteligência artificial (doravante também referida como IA) e suas potencialidades de fazer um mundo melhor. Ou pior. Ou até de aniquilá-lo[6].

Talvez nenhum tema na história da civilização tenha despertado tanta reflexão simultânea. Nos meios de comunicação, nos bares, nas universidades, nos grandes eventos internacionais, nos encontros de especialistas, um assunto se tornou onipresente: a inteligência artificial. Não há aspecto de suas implicações que não venha sendo explorado pelas mentes mais brilhantes e pelos cidadãos mais comuns. O texto que se segue se insere nessa profusão de escritos que procuram captar o espírito do tempo, traçar rotas e empurrar a história na direção certa. Desviando-se dos abismos que colocariam em risco, quando não nossas vidas, pelo menos nossa humanidade, como a conhecemos. A fé na ciência, como toda crença nesse mundo, não pode levar ao fanatismo. Precisamos definir rumos e limites. Aqui segue só mais uma tentativa de fazê-lo.

O presente artigo está estruturado da forma seguinte. Uma introdução apresenta algumas noções básicas acerca do tema. A Parte I explora as potencialidades positivas da IA. A Parte II procura catalogar os principais riscos que a acompanham. A Parte III identifica alguns princípios que devem reger a regulação da matéria. E, ao final, uma conclusão procura aplacar nossas aflições quanto ao futuro.

6 A pesquisa em IA evita o alarmismo, o que não quer dizer que desdobramentos potencialmente catastróficos não sejam considerados como hipóteses sérias. Confira-se sobre o tema a obra de Stuart Russell, *Inteligência artificial a nosso favor. Como manter o controle sobre a tecnologia*. As preocupações mais relevantes orbitam em torno da inteligência artificial geral, também referida como superinteligência artificial, que define um estado em que os computadores superariam as capacidades humanas em uma medida proeminente, acarretando, conforme Kai-Fu Lee (2019, p. 159-173), "problemas de controle" e "problemas de alinhamento".

Introdução
O alvorecer da quarta revolução industrial

1. Um admirável mundo novo

Uma nova revolução industrial desponta no horizonte. A primeira ocorreu em meados do século XVIII e é representada pelo uso do vapor como fonte de energia. A segunda revolução industrial, na virada do século XIX para o século XX, tem como símbolos a eletricidade e o motor de combustão interna. A terceira se desenrolou nas décadas finais do século XX, tendo se consumado com a substituição da tecnologia analógica pela digital. Conhecida como Revolução Tecnológica ou Revolução Digital, ela permitiu a universalização dos computadores pessoais, dos telefones inteligentes e é simbolizada pela internet, conectando bilhões de pessoas em todo o mundo (Barroso, 2019, p. 1262). A quarta revolução industrial, que começa a invadir nossas vidas, vem com a combinação da inteligência artificial, da biotecnologia e a expansão do uso da internet, criando um ecossistema de interconexão que abrange pessoas, objetos e mesmo animais de estimação, numa internet de coisas e de sentidos.

Nesse desafiador mundo novo que se descortina, as novas tecnologias podem nos liberar das atividades mais simples do dia a dia, assim como desempenhar tarefas altamente complexas. Podem limpar ambientes, regular a temperatura, e, em breve, dirigirão carros de forma autônoma (Manyika, 2022, p. 12). Prometem recuperar movimentos corporais perdidos (Caczan, 2023), prover diagnósticos médicos mais precisos (Dilsizian; Siegel, 2014, p. 441), suprir deficiências neurológicas, ampliar habilidades cognitivas (Schmidt, 2017, p. 6-10), criar o "gêmeo virtual" de alguém[7], reproduzir uma pessoa que já morreu[8], permitir o reencontro com entes queridos que já se foram (Here.After [20--]), cuidar de idosos (Horowitz, 2023), encontrar o amigo ou par romântico ideal (Inner Circle, [20--]; Tinder, [20--]), redigir textos nas mais diversas línguas (ChatGPT, [20--]), distribuir auxílios assistenciais aos mais vulneráveis, direcionar serviços públicos de primeira necessidade aos lugares mais carentes (Katyal, 2022, p. 327; Urueña, 2023). Pretendem, ainda, prever a prática ou reincidência de crimes (Eubanks *apud* Eubanks, 2015), melhorar o monitoramento ambiental, promover o planejamento de cidades inteligentes

7 O *site* convida o usuário a "duplicar-se virtualmente", a fim de "ampliar sua produtividade, saúde mental e longevidade" (Mindbank, 2024).

8 O objetivo do sistema é replicar a personalidade dos usuários, sua forma de pensar, falar e outras características, de forma a que possa até mesmo interagir com entes queridos, após a morte da pessoa duplicada (Ramirez, 2023).

(Galaz *et al*, 2021, p. 2), estimar o desempenho de candidatos a um emprego, a probabilidade de adimplemento de um financiamento, bem como o desenvolvimento de doenças graves (Silberg; Manyika, 2019), entre outras questões[9].

Há mais. Estima-se que as mesmas tecnologias possam revelar a orientação sexual de uma pessoa (Morrison, 2021), antever e denunciar a intenção de realizar um aborto (Cox, 2022), substituir centenas de figurantes e atores em Hollywood (Beckett, 2023), criar ou suprimir milhares de postos de trabalho mecânicos ou criativos (Manyika, 2022, p. 20), manipular ou falsear informações, sons, imagens, crenças e vontades (Hacker; Engel; Mauer, 2023, p. 1 e 2), gerar vícios (Mohammad; Jan; Alsaedi, 2023; Becket; Paul, 2024), interferir sobre comportamentos de consumo (Makhnoumi, 2024), influenciar o resultado de processos eleitorais (Heawood, 2018, p. 429-434; Berghel, 2018, p. 84-89), provocar comportamentos violentos (Pauwels, 2020), fortalecer agendas extremistas (Vlachos, 2022), agravar a desigualdade e a discriminação de grupos minoritários (Angwin *et al*., 2016; Eubanks *apud* Eubanks, 2015), alterar e adquirir vontade própria (Hutson, 2023)[10], acionar armas de destruição em massa, colocar a vida, a saúde e a segurança das pessoas em risco (Manyika, 2022, p. 21 e 27).

A lista é interminável e pode nos conduzir ao sublime e ao horror, à liberdade ou à escravidão. À ampla afirmação dos direitos humanos ou à sua supressão. Como intuitivo, o problema não está na tecnologia em si, mas no uso que faremos dela e, sobretudo, em como pretendemos distribuir os benefícios que irá gerar. O desafio, portanto, está em se produzir um desenho institucional que incentive o bom uso da Inteligência Artificial e que contenha o seu desvirtuamento, impedindo a automação da produção de injustiças (Degli-Esposti, 2023, p. 10) e a multiplicação dos riscos existentes (Coeckelbergh, 2023, p. 167).

9 As eventualidades positivas são, de fato, impressionantes, o que leva a que algumas correntes cogitem o uso das novas tecnologias para a transformação dos mecanismos de governança, em favor da instauração de uma "democracia algorítmica", supostamente neutra e eficaz. Não obstante, a neutralidade algorítmica não existe, e a legitimidade democrática se relaciona necessariamene com a representação fundada na vontade popular. Nessa linha, a Assembleia Parlamentar do Conselho da Europa entende que a definição dos objetivos políticos e sociais não pode ser relegada aos algoritmos. Pelo contrário, deve permanecer nas mãos de seres humanos que se submetem a um sistema de responsabilização política e legal (União Europeia, 2022).

10 Teme-se que a capacidade de aprendizado autônomo da IA possa levá-la a adquirir inteligência super humana, tornando-a incontrolável. A tal fenômeno designa-se "singularidade".

2. O que é a inteligência artificial

Numa definição simples, é possível afirmar que a inteligência artificial consiste em programas (*softwares*) que transferem capacidades humanas para computadores. Essas capacidades incluem tarefas cognitivas e tomada de decisões, via de regra com base nos dados, instruções e objetivos com que são alimentados[11]. Não há, contudo, uma convergência plena sobre o conceito técnico de IA e sua abrangência[12]. Inúmeras entidades e instituições, como a OCDE[13] e a Unesco[14], procuram delimitar-lhes os contornos. É possível apontarem-se alguns traços comuns a essas tentativas de definição: são sistemas com capacidade de processar dados e informações de forma assemelhada à inteligência humana, que incluem aprendizado, raciocínio, percepção e comunicação por via de linguagem. Consultado, o ChatGPT4 forneceu a seguinte definição:

> Inteligência Artificial (IA) é um ramo da ciência da computação dedicado a criar sistemas capazes de realizar tarefas que, tradicionalmente, requerem inteligência humana. Estas tarefas incluem aprendizado (a capacidade de melhorar o desempenho com a experiência), raciocínio (a capacidade de resolver problemas através de métodos lógicos), percepção (a capacidade de interpretar dados sensoriais para entender aspectos do mundo) e interação linguística (a capacidade de compreender e produzir linguagem natural).

11 A expressão "inteligência artificial" é atribuída a um *workshop* realizado em 1956, em Dartmouth, com o objetivo de se buscar desenvolver máquinas capazes de solucionar problemas resolvidos por humanos e se aprimorarem (MacCarthy *et al*., 1955; Manyika, 2022, p. 15).

12 Organizações representativas de empresas de IA postulam a formulação de um conceito mais restritivo de inteligência artificial, ao passo que organizações de defesas de direitos humanos defendem a ampliação do conceito para outras tecnologias, que também podem produzir efeitos adversos sobre direitos humanos. Em tal contexto, a própria abrangência do contexto de IA depende, em parte, do quanto se pretende regular (Madiega, 2023, p. 6-8).

13 "Um sistema de IA é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir das informações que recebe, como gerar resultados como previsões, conteúdos, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam nos seus níveis de autonomia e adaptabilidade após a implantação" (OCDE, 2019; Russel; Perset, Marko, 2023).

14 "Por conseguinte, a presente recomendação aborda os sistemas de IA como sistemas que têm a capacidade de processar dados e informações de uma forma que se assemelha a um comportamento inteligente e que, normalmente, inclui aspectos de raciocínio, aprendizagem, perceção, previsão, planeamento ou controle" ("*Therefore, this Recommendation approaches AI systems as systems which have the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control*") (Unesco 2021).

No estágio atual (Degli-Esposti, 2023, p. 10; Rebollo Delgado, 2023, p. 24)[15], a inteligência artificial não tem consciência de si mesma, não tem discernimento do que é certo ou errado, tampouco possui emoções, sentimentos, moralidade ou mesmo senso comum. Vale dizer: ela é inteiramente dependente da inteligência humana para alimentá-la, inclusive com valores éticos. Computadores não têm vontade própria (Winston, 2018; Lenharo, 2023). Embora esse seja o conhecimento convencional na matéria, algumas experiências revelam surpreendente capacidade de aprendizado, suscitando novas preocupações. Uma delas foi o Alpha Zero, um programa de IA desenvolvido pela Google que derrotou o Stockfish, até então o mais poderoso programa de xadrez no mundo. Ao contrário de programas anteriores, Alpha Zero não foi alimentado com movimentos previamente concebidos pelo homem. Ou seja: não se baseou no conhecimento, na experiência e nas estratégias humanas. Ele recebeu apenas as regras do jogo. Alpha Zero treinou jogando consigo mesmo, desenvolveu os seus próprios movimentos e estratégias, originais e inortodoxas, com uma lógica própria (Kissinger; Schmidt; Huttenlocher, 2021, p. 7 *et seqs*. e 26).

Duas visões disputaram a primazia nas pesquisas sobre inteligência artificial ao longo dos anos. A primeira delas inspirou-se no modo de funcionamento da mente humana, procurando mimetizar a maneira como elaboramos as questões e desenvolvemos raciocínios lógicos. Essa primeira perspectiva dominou as experiências sobre IA até a década de 80 do século passado. A segunda visão inspirou-se no modo de funcionamento das estruturas do cérebro humano. Propôs, assim, conectar unidades de processamento de informações, equivalentes a neurônios, de modo a simular como eles funcionam (Dreyfus; Dreyfus, 1988, p. 15-44). Essa a visão que se tornou dominante no cenário da IA, denominada "abordagem conexionista" (*connectionist approach*). Ela não procura reproduzir a forma de se racionalizar da mente humana. Ao contrário, busca estabelecer correlações e padrões entre milhares de dados e determinados resultados. Seus principais pontos de apoio são a estatística e a neurociência.

Os sistemas de Inteligência Artificial baseiam-se em dados e algoritmos. Quanto maior o conjunto de dados a que têm acesso, maior é o número de correlações confirmadas e descartadas e, naturalmente, mais precisos tendem a ser os resultados (Dreyfus; Dreyfus, 1988, p. 15-44). Um determinado

15 A ressalva se impõe tendo em vista que não se descarta que a IA do futuro conceda, às máquinas, doses intensas de autonomia e de consciência, em um panorama em que as aplicações inteligentes adquiram uma racionalidade própria, perseguindo objetivos não previstos.

universo de dados ou características correlacionadas leva a IA a identificar um cachorro ou um gato, um bom ou um mau devedor, uma pessoa com tendências depressivas, uma criança em risco. O estabelecimento de correlações entre tais elementos pode parecer aleatório ou irracional para o modo de conhecer da mente humana. Mas, relembre-se, o modelo é baseado em estatística, não em lógica.

Algoritmo, por sua vez, é um conceito fundamental em ciência da computação. O termo identifica o conjunto de instruções, regras e parâmetros que orientam os computadores a cumprir as tarefas que lhes foram atribuídas. São fórmulas, códigos e roteiros que selecionam, tratam e estocam os dados, com o objetivo de obter um determinado resultado. Os dados selecionados (*inputs*) e suas correlações permitem conduzir aos resultados visados pelo programa (*outputs*), que podem ser os mais variados. Por exemplo: se o resultado dá ensejo à diferenciação entre objetos e seres vivos, fala-se em IA discriminativa; se o resultado for a previsão de comportamentos – de consumo, financeiros ou políticos – tem-se a IA preditiva; se for a geração de conteúdos – textos, imagens ou sons –, diz-se que é IA generativa (Hacker; Engel; Mauer, 2023, p. 1-3 e 13)[16].

3. Aprendizado de máquina, modelo fundacional e outros conceitos relevantes

No que se refere ao modo de operar, os sistemas de inteligência artificial mais avançados atualmente são aqueles capazes de desenvolver o aprendizado de máquina. O aprendizado de máquina refere-se à aptidão de um modelo para adquirir conhecimento autonomamente, sem prévia programação explícita, com base na identificação de correlações entre grandes volumes de dados, como descrito acima. Vale observar, ainda, que, para conceitos mais restritos de IA, a capacidade de aprendizado de máquina é o que diferencia a inteligência artificial da mera automação, que seria um fenômeno mais amplo (Nunes; Andrade, 2023, p. 4; Brown, 2021). O aprendizado de máquina é o processo que serve de base a grande parte dos serviços de IA que usamos hoje, tais como os sistemas de recomendação de conteúdos de plataformas como Netflix, YouTube e Spotify, os modelos de seleção e hierarquização de resultados em ferramentas de busca como Google, Bing e Baidu, além de *feeds* e regimes de recomendação de contatos em mídias sociais como Facebook e X (ex-Twitter) (Hao, 2018; Nunes; Andrade, 2023, p. 5).

16 As qualificações acima não são exaustivas. Alude-se, por exemplo, à IA adversarial, destinada a impedir o funcionamento de outra IA para fins de proteção de dados, entre outras (Hacker; Engel; Mauer, 2023, p. 13).

O aprendizado de máquina se vale dos algoritmos e das redes neurais artificiais. As "redes neurais" artificiais (*neural networks*) se inspiram em redes de neurônios humanos. São modelos matemáticos que imitam nosso sistema nervoso (Porto; Araújo; Gabriel, 2024, p. 37). Por meio delas, diferentes processadores de dados trabalham conjuntamente para tratar tais dados. O "aprendizado profundo" de máquina (*machine deep learning*) é a técnica que amplia a capacidade de aprendizado de máquina, por meio do estabelecimento de várias camadas de redes neurais artificiais, simulando o complexo funcionamento do cérebro humano, a fim de que tais múltiplas redes e processadores administrem as informações e estabeleçam correlações simultaneamente (Hao, 2018; Re; Solow-Niederman, 2019, p. 244-246). Tem aplicação no reconhecimento de imagem e fala, tradução automática e processamento de texto. Existem três espécies de aprendizado de máquina: supervisionado (*supervised*), não supervisionado (*unsupervised*) e por reiteração (*reinforcement*) (Brown, 2021)[17].

Quanto ao uso ou à finalidade, os sistemas de IA comportam inúmeras classificações, que por vezes se superpõem. Modelos fundacionais (*foundational models*) são treinados com grandes quantidades de dados, preparados para se adaptarem a múltiplas tarefas. O ChatGPT (*Chat Generative Pre-trained Transformer*), uma IA generativa, é exemplo de modelo fundacional (também chamado de "propósito geral") e de um "grande modelo de linguagem" (*large language model*), com capacidade para gerar arte, imagens, textos e sons (Jones, 2023)[18]. Note-se bem: programas generativos de IA podem criar conteúdos novos, não apenas analisar ou classificar conteúdos existentes. Se alguém pesquisar no Google como funciona um carro elétrico, será remetido a um *link* que levará a um *site* de terceiros. O ChatGPT, no entanto, irá explicar como um carro elétrico funciona nas suas próprias palavras (Ariyaratne, 2023, p. 4 e 5).

Os modelos fundacionais são considerados um passo na direção da chamada IA de propósito geral (*general purpose AI*), que ainda não foi inteiramente alcançada, mas já é capaz de realizar uma grande quantidade de tarefas, não

17 O aprendizado supervisionado, que é o principal tipo utilizado atualmente, é aquele em que é possível definir-se qual é o *output* adequado. É o caso dos sistemas de IA voltados a categorizações (em gatos, cachorros, meios de transporte). O aprendizado não supervisado (*unsupervised*), menos comum, é aquele que busca a identificação de padrões, correlações e agrupamentos sem a prévia definição do resultado adequado. Por fim, o aprendizado por reiteração (*reinforcement*) tem o objetivo de treinar e aprimorar máquinas por meio de um sistema de tentativa e erro, de modo a ensiná-las a tomar as melhores decisões (Brown, 2021).

18 Nem toda IA generativa constitui um modelo fundacional. Ela pode ser modelada para propósitos bem específicos. As capacidades generativas podem incluir a manipulação e a análise de texto, de imagem, de vídeo e a produção de discurso. Aplicativos generativos incluem *chatbots*, filtros de fotos e vídeos e assistentes virtuais.

sendo limitada a objetivos específicos (Bommasani *et al*., 2022, p. 4-12; Hacker; Engel; Mauer, 2023, p. 2-4; Uuk; Gutierrez; Tamkin, 2023). Por sua vez, os modelos de propósito determinado ou estrito (*fixed-purpose AI systems* ou *narrow AI*), como o próprio nome sugere, destinam-se a um propósito específico e, por isso, são conformados de forma mais restritiva. São treinados com uma base mais direcionada de dados. Nessa categoria enquadra-se a maior parte dos sistemas de IA atualmente em uso, que incluem os assistentes de voz como Siri, Alexa e Google Assistant, que se prestam a compreender e cumprir comandos de voz; sistemas de recomendação de conteúdo em plataformas de *streaming*, filtros de *spam*, sistemas de predição do tempo, *softwares* de reconhecimento facial, entre outros. Todos se destinam a uma finalidade bastante delimitada.

Por fim, a expressão IA forte (*strong AI*), também referida como inteligência artificial geral (*AGI – artificial general intelligence*), designa sistemas com capacidade de compreensão, aprendizado e aplicação concreta equivalente à dos seres humanos. Seria capaz, assim, de raciocínio, resolução de problemas e tomadas de decisões próprias. Esse tipo de IA constitui um conceito teórico, ainda não alcançado, mas possível de ser concretizado em alguns anos, segundo alguns pesquisadores (Taulli, 2020, p. 218).

Seria possível continuar explorando as múltiplas tecnicalidades do tema. Como, por exemplo, a cadeia de valor da IA, com suas diferentes fases: *design*, desenvolvimento, implementação e manutenção do sistema (Hacker; Engel; Mauer, 2023, p. 8-11). Cada uma dessas fases tem seus próprios desafios e riscos, envolvendo diferentes atores, que incluem o desenvolvedor (*developer*)[19], o implementador (*deployer*)[20], o usuário[21] e o destinatário final (*recipient*)[22]. Tais agentes detêm distintas *expertises* e habilidades, podendo gerar diferentes aportes, danos e responsabilidades (Hacker; Engel; Mauer, 2023, p. 8-11). Essa variedade de papéis, como intuitivo, acrescenta alguns graus de dificuldade à regulação da matéria.

19 Seria o caso da OpenAI quanto ao GPT-4 (modelo fundacional).

20 No caso do ChatGPT (modelo de propósito específico), a OpenIA é desenvolvedora e implementadora. A Be My Eyes, a seu turno, é apenas implementadora do Virtual Volunteer (também de propósito específico), adaptado a partir do GPT-4.

21 No caso dos produtos referidos na nota anterior, usuário é quem gera o texto com o ChatGPT ou consulta o aplicativo de voz, via Virtual Volunteer.

22 Quem usa o texto e as orientações de voz. Veja que a notícia produzida pode ou não ser verdadeira. A voz pode ser uma boa orientação ou uma *deep fake*, produzida para iludir o destinatário.

É hora, todavia, de seguir adiante, explorando implicações da inteligência artificial que transcendem os temas estritamente técnicos.

PARTE 1
A INTELIGÊNCIA ARTIFICIAL E SEUS BENEFÍCIOS

A inteligência artificial vem crescentemente se incorporando à rotina das nossas vidas, por vezes de maneira tão natural que nem associamos certas utilidades a ela. A verdade é que a vida analógica vai ficando para trás. É certo que sempre haverá quem prefira objetos manufaturados ao estilo antigo, como alguns relógios de marcas caríssimas, que celebram um passado artesanal, embora atrasem, adiantem e funcionem bem pior que seus equivalentes digitais. Mas continuam atraindo compradores, a comprovar que a espécie humana não é totalmente movida pela razão e pelo pragmatismo. Porém, salvo extravagâncias e idiossincrasias, o fato é que hoje fazemos pesquisas sobre qualquer tema por meio de algoritmos de busca. Escolhemos produtos, leituras, viagens, hospedagens utilizando algoritmos de recomendação. Optamos por deslocamentos mais rápidos ou definimos o que vestir com base em sistemas inteligentes de medição de tráfego de veículos e de medição de temperatura. Em suma, a IA traz muitas coisas positivas, que tornam nossa vida melhor e mais fácil.

Pensando no impacto em larga escala da inteligência artificial, são tantas as suas utilidades e potencialidades que não é sequer fácil selecioná-las e sistematizá-las. A seguir, alguns exemplos significativos.

1. Melhor capacidade decisória em muitas áreas

Em inúmeros domínios, a IA terá melhor capacidade de tomada de decisões que seres humanos, por variadas razões. Em primeiro lugar, por poder armazenar uma quantidade de informações bem maior do que o cérebro humano. Em segundo lugar, por ser capaz de processá-las com muito maior velocidade. Em terceiro lugar por ser capaz de fazer correlações dentro de um volume massivo de dados, para além das possibilidades de uma pessoa ou mesmo de uma equipe. Tais correlações podem revelar associações entre fatores dos quais não nos damos conta, por sua complexidade ou sutileza. Como já assinalado, no entanto, a eficiência da IA dependerá da quantidade e da qualidade dos dados com que alimentada. Ademais, no atual estado da arte,

ferramentas de IA generativa podem produzir informações inventadas ou absurdas, num desvio conhecido como "alucinação" (*hallucination*)[23]. Deve-se ressaltar que em áreas que dependam de inteligência emocional, de valores éticos ou de compreensão das nuances do comportamento das pessoas, a intervenção humana será indispensável e sua capacidade decisória superior.

2. Automação

A IA permite a automação de inúmeras tarefas, tanto rotineiras como mais complexas, aumentando a produtividade e a eficiência em várias áreas de atividade. Tarefas repetitivas, desgastantes ou extenuantes para pessoas humanas podem ser desempenhadas por máquinas, como, por exemplo, em linhas de produção industrial. Além disso, reduz-se a margem de erros e é possível a eliminação de riscos, em trabalhos como exploração de minas, desarme de bombas, reparo de cabos no fundo do oceano ou viagens espaciais. Acrescente-se que a IA pode trabalhar ininterruptamente por 24 horas, todos os dias da semana, produzindo em maior escala, com melhor precisão e a menor custo. Não cansa, não adoece, não varia de humores e não há risco de ajuizar reclamação trabalhista. O impacto negativo que tudo isso pode produzir sobre o mercado de trabalho será examinado adiante.

3. Linguagem

O impacto da IA sobre o campo da linguagem tem sido profundo e multifacetado, especialmente pelo uso do Processamento de Linguagem Natural (*Natural Language Processing*). A qualidade das traduções feitas pelo Google Translate, pelo ChatGPT e pelo DeepL, para citar alguns exemplos, foi aprimorada de maneira expressiva, tornando-as bastante precisas e fluentes. Com isso, rompem-se muitas das barreiras do idioma na comunicação humana. Ferramentas como Siri, Alexa e Google Assistant respondem a comandos de voz. Outras ferramentas transformam textos em fala. Chatbots ajudam a resolver dúvidas e problemas de consumidores e clientes. E a IA generativa, que vem assombrando o mundo, comunica-se com o usuário por meio de textos, sons e imagens. São extraordinários os avanços nessa área.

23 Note-se que as alucinações de IA podem repercutir de forma brutal em processos sociais relevantes. Por exemplo, de acordo com um relatório do laboratório *AI Forensics,* o *chatbot* Bing, da Microsoft, ofereceu informações equivocadas em 30% das consultas básicas que lhes foram feitas sobre temas afetos a eleições na Alemanha e na Suíça. Descobriu-se, ainda, que o problema crescia quando as perguntas eram feitas em idiomas diferentes do inglês (Guadián, 2024).

4. Pesquisa e inovação

A IA ampliou as fronteiras da pesquisa e da inovação em quase todas as áreas da atividade humana, da física e da química à indústria automobilística e espacial. O volume de ciência que se vem produzindo com base na IA tem crescido exponencialmente. A IA pode simplificar e abreviar pesquisas clínicas e testes com novos medicamentos, materiais e produtos. A análise de vastas quantidades de dados acelera o processo de descobertas científicas. Importante destacar a redução de custo e de tempo no desenvolvimento de novas drogas, bem como de carros autônomos, com a promessa de redução do número de acidentes. A expectativa é que a IA, na sua relação com a pesquisa e a inovação, possa ajudar no enfrentamento de inúmeros desafios da humanidade, como a mudança climática, o combate à fome, o controle de pandemias, a sustentabilidade das cidades e doenças como câncer e alzheimer (União Europeia, 2023a). O movimento que promove a exploração benéfica dos potenciais da IA é conhecido como *Data for good* (Muñoz Vela, 2022, p. 65).

5. Aplicações na medicina

A medicina é uma das áreas de maior impacto da inteligência artificial sobre a vida e a saúde das pessoas. Tecnologias como o aprendizado de máquina e o processamento de linguagem natural vão melhorar a qualidade da assistência aos pacientes e reduzir custos. O aperfeiçoamento de diagnósticos, a análise de imagens, as cirurgias robóticas, o planejamento e a personalização dos tratamentos, a telemedicina, a previsão de futuras doenças e o manejo dos dados dos pacientes são alguns dos múltiplos benefícios que poderão advir. A IA não tornará a atuação do médico prescindível, mas pode modificar alguns dos papéis que desempenha, transferindo atribuições do plano técnico para o plano humano da empatia e da motivação. Também haverá implicações éticas e legais, como, por exemplo, erros praticados por equipamentos de IA (Davenport; Kalacota, 2019,).

6. Aplicações no sistema de justiça

A IA traz a perspectiva de transformações profundas na prática do direito e na prestação jurisdicional. Num ambiente em que os precedentes vão se tornando mais importantes, é enorme a sua valia para a pesquisa eficiente de jurisprudência. A possibilidade de elaboração de peças por advogados, pareceres pelo Ministério Público e decisões pelos juízes, com base em minutas

pesquisadas e elaboradas por IA irá simplificar a vida e abreviar prazos de tramitação. Por evidente, tudo sob estrita supervisão humana, pois a responsabilidade continua a ser de cada um desses profissionais. Nos tribunais, programas de IA que agrupam processos por assuntos, bem como os que podem fazer resumos de processos volumosos otimizam o tempo e a energia dos julgadores. Da mesma forma, a digitalização dos processos – no Brasil, hoje, a quase totalidade dos processos e de sua tramitação são eletrônicos –, a automação de determinados procedimentos e a resolução *online* de conflitos têm o potencial de tornar a Justiça mais ágil e eficiente. No Brasil, no âmbito dos diversos tribunais do país, existe mais de uma centena de projetos de utilização da IA na prestação jurisdicional.

Há aqui um ponto controvertido e particularmente interessante: o uso da IA para apoiar a elaboração de decisões judiciais. Muitos temem, não sem razão, os riscos de preconceito, discriminação, falta de transparência e de explicabilidade. Sem mencionar ausência de sensibilidade social, empatia e compaixão. Mas é preciso não esquecer que juízes humanos também estão sujeitos a esses mesmos riscos. Por essa razão, há um outro lado para essa moeda: a perspectiva de que a IA possa, efetivamente, ser mais preparada, imparcial e menos sujeita a interesses pessoais, influências políticas ou intimidações. Isso pode acontecer em qualquer lugar, mas especialmente em países menos desenvolvidos, com menor grau de independência judicial ou maior grau de corrupção (Ariel Gustavo, 2021; Sustein, 2022). Seja como for, no atual estágio da civilização e da tecnologia, a supervisão de um juiz humano é indispensável, embora se possa impor a ele um ônus argumentativo aumentado nos casos em que pretenda produzir resultado diverso do proposto pela IA.

7. Educação e cultura

A inteligência artificial vai transformar a paisagem da educação no mundo, tanto nos métodos de ensino quanto nas possibilidades de aprendizado. De início, a internet, potencializada pela IA, ampliou exponencialmente o acesso ao conhecimento e à informação, expandindo o horizonte de todas as pessoas que têm acesso à rede mundial de computadores. Além disso, a educação a distância rompeu as barreiras de tempo e espaço, permitindo o aprendizado a qualquer hora, de qualquer lugar do mundo. Bibliotecas digitais dispensam o deslocamento físico e permitem a consulta em repertórios situados em qualquer lugar do mundo. Na perspectiva dos professores, ela pode ajudar na preparação de aulas, na elaboração de questões e mesmo na correção de

trabalhos, além de desempenhar tarefas administrativas que liberam os professores para mais tempo de atividade acadêmica. Tudo, sempre, reitere-se, com supervisão humana.

Do ponto de vista dos alunos, a IA, sobretudo generativa, facilita a pesquisa, pode resumir textos longos, corrigir erros gramaticais e sugerir aperfeiçoamentos de redação, além de ajudar a superar as barreiras linguísticas, como visto acima. Ela permite, também, a personalização do ensino, customizado às necessidades dos alunos, inclusive para pessoas com deficiência, na medida em que, por exemplo, pode transformar texto em voz ou vice-versa[24]. Naturalmente, para trazer benefícios distribuídos de maneira equânime pela população, o uso da IA pressupõe conectividade de qualidade para todos (inclusão digital). Não se deve descartar, aqui, algumas disfunções que podem advir do uso da IA na educação, que vão do plágio à limitação da criatividade e do espírito crítico. Por isso mesmo, organizações internacionais como a OCDE (2023) e a Unesco (2021) produziram documentos relevantes, com princípios e diretrizes para o uso da IA na educação.

O impacto da IA sobre a cultura também será imenso. Na face positiva, ela abrirá caminhos para a criatividade, em sinergia com músicos, pintores, escritores, arquitetos, desenhistas gráficos e em inúmeros outros agentes de criação. A IA generativa pode auxiliar na composição de sinfonias, obras literárias, poesias, narrativa de histórias etc., aumentando a criatividade, o universo estético, mas também suscitando inúmeras questões de natureza ética sobre propriedade intelectual e direitos autorais (Beiguelman, 2021, p. 59). Merece reflexão a afirmação de Yuval Noah Harari de que a IA já *hackeou* o sistema operacional da cultura humana, que é a linguagem. E indaga: o que significará para os seres humanos viver num mundo em que um percentual dos romances, músicas, imagens e leis, em meio a muitas outras criações, são gerados por uma inteligência não humana? (Harari, 2023)

8. Outras aplicações úteis da IA

8.1 Utilidades práticas do dia a dia

A tecnologia da IA está presente nos computadores pessoais e nos telefones celulares inteligentes em múltiplos aplicativos, como Google Maps, Waze,

24 Na área educacional, a IA pode beneficiar ainda os estudantes portadores de altas capacidades (superdotados), dado que as habilidades de percepção, reconhecimento e recomendação permitem supervisionar, compreender e adaptar o processo de aprendizado de cada aluno, além de liberar os professores para um tempo maior para a instrução individual (Lee, 2019, p. 149).

Uber, Spotify, Zoom, Facebook, Instagram. E, também, em assistentes pessoais, como Siri e Alexa. A IA tem papel importante, igualmente, na indústria de entretenimento via *streaming* (Neflix, Amazon Prime, HBO Max) e de *games*. Sem mencionar os aplicativos que permitem transações bancárias e pagamentos por cartões de crédito, em meio a inúmeras outras utilidades.

8.2 Proteção do meio ambiente

A IA terá um papel cada vez mais crítico em relação à proteção ambiental, na análise de dados, na previsão de fenômenos e no monitoramento de situações. Os exemplos são múltiplos e incluem: exame de dados sobre mudança climática, uso de imagens de satélites e *drones*, controle dos níveis de poluição do ar, da água e do solo, racionalização da distribuição e do consumo de energia e de água, previsão de desastres naturais (como furacões, terremotos e inundações), auxílio na agricultura sustentável por meio de sensores de solo e outros instrumentos, com redução do uso de pesticidas, orientação à irrigação e ajuda no planejamento do reflorestamento[25].

8.3 Personalização das relações comerciais e outras

A IA permite que indústria, comércio, serviços, meios de comunicação e plataformas digitais direcionem a seus consumidores informações, notícias e anúncios que correspondam aos seus interesses. Isso, naturalmente, otimiza o tempo das pessoas e facilita a aquisição de produtos, livros, planejamentos de viagem e inúmeras outras escolhas a serem feitas e decisões que precisam ser tomadas. Recomendações de filmes, de músicas ou de outras formas de entretenimento vêm desse uso da inteligência artificial. Não se deve desconsiderar aqui, todavia, aspectos negativos associados a uma certa tribalização da vida, pelo viés de confirmação decorrente do envio de materiais que, no geral, reiteram preferências e convicções. Tal fenômeno reduz a pluralidade de visões, gera novas formas de controle social[26] e pode conduzir à polariza-

25 No entanto, o gasto de energia decorrente da alimentação, operação e manutenção da IA, bem como seus impactos sistêmicos sobre diferentes ecossistemas tampouco devem ser ignorados, como alguns pesquisadores já observaram (García-Martín *et al*., 2019, p. 75-88; Centeno *et al*., 2021, p. 1-10).

26 Sobre o uso da IA como forma de controle social: "Os mecanismos de busca apresentam um outro desafio: dez anos atrás, quando eram movidos por *data mining* (em vez de aprendizado de máquina), se uma pessoa fizesse buscas por um "restaurante gourmet", e depois por "roupas", sua última busca seria independente da primeira. Nas duas vezes, um mecanismo de pesquisa agregaria o máximo de informações possível e lhe daria opções [...]. As ferramentas contemporâneas, por sua vez, são guiadas pelo comportamento humano observado. [...]. A pessoa pode estar procurando roupas de grife. No entanto, existe uma diferença entre escolher em meio a uma variedade de opções e realizar uma ação – nesse caso, fazer uma compra; em outros casos, adotar uma posição ou uma ideologia política [...] – sem nunca ter visto o leque inicial de possibilidades ou implicações, apenas confiando em uma máquina para configurar antecipadamente as opções" (Kissinger; Schmidt; Huttenlocher, 2023, p. 20).

ção e ao radicalismo (Barroso; Barroso, 2023). No plano das relações pessoais, pesquisas demonstram que casamentos resultantes de relacionamentos iniciados *online*, com auxílio de algoritmos, têm se revelado ligeiramente mais satisfatórios que os casamentos em que os parceiros se conhecem por métodos convencionais, *offline* (Tropiano, 2023; Harms, 2013).

Não é o caso de se seguir listando, indefinidamente, todas as utilidades e os benefícios decorrentes da inteligência artificial, que, de resto, se ampliam a cada dia. Entre eles se incluem o desenvolvimento dos veículos autônomos, o monitoramento de equipamentos para detectar possíveis falhas em infraestruturas, a detecção de fraudes, sobretudo de natureza financeira, o aprimoramento da cybersegurança, os controles de aviação etc. Cabe, agora, voltar os olhos para os problemas, riscos e ameaças que podem decorrer da utilização em larga escala da inteligência artificial.

PARTE II
A INTELIGÊNCIA ARTIFICIAL E SEUS RISCOS

Toda nova tecnologia produz um efeito disruptivo sobre as relações de produção, de consumo e sobre o mercado de trabalho, impactando a vida social. Além disso, como muitas coisas na vida, as inovações podem ter um lado negativo ou ser apropriadas por maus atores sociais. A máquina de tear desempregou costureiras e artesãos; a impressão em *offset* eliminou os empregos de linotipista. A informatização diminuiu a necessidade de bancários no sistema financeiro. As plataformas digitais abriram caminho para a polarização extremista (Fisher, 2023, p. 20), a desinformação (Kakutani, 2018, p. 17) e os discursos de ódio (Campos Mello, 2020; Williams, 2021, p. 207). Mais grave ainda: a invenção das caravelas permitiu o comércio transoceânico, mas também o tráfico negreiro (Acemoglu; Simon; 2023, p. 4-5)[27].

Por essas razões, é preciso ter atenção para os efeitos adversos do uso da inteligência artificial, procurando neutralizá-los ou mitigá-los. Tais impactos negativos da IA podem ter implicações sociais, econômicas, políticas ou até mesmo abalar a paz mundial. A seguir, o levantamento de algumas consequências, riscos e ameaças trazidas pela inteligência artificial.

27 Os autores apontam algumas invenções que, nos últimos mil anos, não trouxeram, necessariamente, prosperidade para todos.

1. Impacto sobre o mercado de trabalho

Esse é o efeito mais óbvio e previsível, fruto do que normalmente ocorre quando uma nova tecnologia abala o modo de produção anterior. Com o avanço da automação, a paisagem do mercado de trabalho irá se modificar profundamente, exigindo adaptação dos trabalhadores de áreas diversas da economia para novos trabalhos. Uma transição que nem sempre é fácil. Note-se que no caso da IA, o impacto será não apenas quanto a postos de trabalhos mais mecânicos, mas afetará, também, funções mais qualificadas e criativas[28]-[29]. É certo que novas tecnologias também tendem a gerar novos mercados e, consequentemente, novos empregos. Entretanto, há um problema de *timing* e de escala nessa consideração. É improvável que novos postos de trabalho sejam espontaneamente gerados no mesmo ritmo e volume (Keynes, 1930)[30]. Esse é um importante desafio, que exigirá dos governos investimento em proteção social e capacitação dos trabalhadores, convindo lembrar que a expansão da vulnerabilidade econômica tende a impactar a esfera de proteção democrática, dado que assoma, historicamente, como um fator potencial de desestabilização.

2. Utilização para fins bélicos

É relativamente escassa a literatura acerca da utilização da IA para fins bélicos, até pelo sigilo que normalmente se impõe na matéria, por motivos de segurança. Mas ao longo da história, novas tecnologias ou são originárias de pesquisas para objetivos militares ou são rapidamente direcionadas a esse fim. Não é difícil imaginar países como Estados Unidos e China numa competição para emprego da IA com destinação militar, com uso das novas tecnologias e de robôs. Aliás, *drones* automatizados (*automated drones*) operados remotamente já são utilizados há algum tempo para esse fim, com missões de reconhecimento, vigilância, entrega de equipamentos ou mesmo ataques aéreos. Tema que tem despertado grande preocupação é o das armas letais

28 Estima-se que os bancos e algumas empresas de tecnologia gastam 60% a 80% das suas folhas de pagamento, ou mais, com trabalhadores com alta probabilidade de serem afetados pela nova tecnologia (Lohr, 2024). Em sentido semelhante, veja-se em Maheshwari (2024) quanto ao mercado de propaganda e *marketing*.

29 Outros estudos indicam que funções que demandem inteligência social (relações públicas), criatividade (biólogos e *designers*), percepção e manipulação fina (cirurgiões) tendem a ser mais poupados (Frey; Ousborne, 2017). Avaliam, ainda, que trabalhos de análise, previsão e estratégia serão os mais afetados (Webb, 2019).

30 Como observado por Keynes (1930), há quase um século, o avanço tecnológico tende a gerar um desajuste ao menos temporário em matéria de trabalho, até que novas oportunidades de trabalho são identificadas.

autônomas (*autonomous lethal weapons*), que podem se engajar em combate e atacar alvos por decisão própria, sem controle humano. Há debates em curso acerca do controle estrito do seu uso por atos internacionais (Klare, 2023). As implicações éticas desse tipo de armamento são dramáticas e é imperativa uma regulação rigorosa do seu uso ou, talvez preferencialmente, o seu banimento.

Além disso, as tecnologias de comunicação e informação já há algum tempo têm mobilizado esforços militares, protagonizando táticas recorrentes no contexto das "guerras híbridas" (*hybrid warfares*). Trata-se de novas formas de agressão que envolvem, além da destruição por meios físicos, campanhas de influência e desinformação (*cognitive warfares*), além de ciberataques com o propósito de comprometer sistemas informatizados vitais, como por exemplo as estruturas de fornecimento de energia (Alvim; Zilio; Carvalho, 2023, p. 69).

3. Massificação da desinformação

Ao menos desde 2016, a difusão de informações por meio de plataformas digitais e aplicativos de mensagens tem representado um problema grave para o processo democrático e eleitoral. Estudos documentam que a circulação de falsidades e o radicalismo *online* se dão em maior velocidade e com maior engajamento do que a difusão de discursos verdadeiros e moderados. O que é emocional, improvável, alarmante produz mais engajamento e mobilização. O *deep fake* torna as coisas ainda piores, na medida em que simula pessoas falando coisas que jamais disseram, adulterando conteúdos e realidades de forma imperceptível para o cidadão (Barroso; Barroso, 2023; Campos Mello; Rudolf, 2023, p. 53-78). Tal panorama não é hipotético e os antecedentes são preocupantes. Tornou-se notória a influência que a disseminação de desinformação exerceu sobre eventos históricos como a saída do Reino Unido da União Europeia (Brexit), as eleições nos Estados Unidos, ambas em 2016, e nas eleições brasileiras de 2018. A democracia pressupõe a participação esclarecida dos cidadãos e, naturalmente, fica gravemente comprometida com a circulação ampla de mentiras deliberadas, destruição de reputações e teorias conspiratórias.

4. Violação da privacidade

O modelo de negócios das plataformas que se valem da IA se baseia na coleta da maior quantidade possível de dados pessoais dos indivíduos, o que transforma a privacidade em mercadoria (Morozov, 2018, p. 36). Com base

neles, algoritmos complexos e múltiplas camadas neurais estabelecem correlações profundas, que permitem obter seus dados genéticos, seus sistemas psíquicos, vulnerabilidades, comportamentos de consumo, políticos, financeiros, sexuais, religiosos (Huq, 2020, p. 37). Com tais dados e correlações, a IA é capaz de realizar predições, recomendações, manipular interesses e produzir os resultados almejados pelo algoritmo. Portanto, o acesso a dados privados, de pessoas e de empresas, é central para o modelo de negócios da IA tal como atualmente estabelecido (Zuboff, 2022, p. 1-79)[31]. Não por acaso, no meio acadêmico, os dados têm sido tratados como o petróleo do século em curso (Rebollo Delgado, 2023, p. 17).

Há pelo menos três aspectos que exigem atenção, relativamente ao tema da privacidade. O primeiro deles é a obtenção de dados dos usuários da internet, sem o seu consentimento, pelas plataformas digitais e *sites* da internet. Tais informações são utilizadas para venda comercial, para direcionamento de informações e publicidade ou mesmo para a manipulação da vontade dos usuários, como pesquisas acerca da neurociência demonstram. Um segundo aspecto diz respeito à vigilância e ao rastreamento pelo governo e por autoridades policiais, mediante tecnologias de reconhecimento facial e ferramentas de localização. Embora o fim legítimo seja o combate à criminalidade, os riscos de abuso são muito grandes. Esses riscos, como intuitivo, se agravam no caso de governos autoritários. Por fim, um terceiro ponto é que os sistemas de IA exigem a obtenção de vastas quantidades de dados para treinar os respectivos modelos, com os riscos de vazamento e ataques cibernéticos por atores maliciosos, por exemplo, em atividades de *spear phishing* (Muñoz Vela, 2022, p. 64)[32] e *doxxing* (Prado, 2023, p. 162)[33] que não raro alimentam práticas de assédio, violência política, *malinformation* e desinformação.

31 Muitas das demais restrições de direitos derivam da restrição à privacidade, tais como aquelas relacionadas a: danos físicos, reputacionais, relacionais, psicológicos (emocionais), econômicos, discriminatórios e relacionados à autonomia humana (coerção, manipulação, desinformação, deformação de expectativas, perda de controle entre outros) (Citron; Solove, 2022, p. 793-863; Huq, 2020).

32 Trata-se de *e-mail* ou mensagens mal intencionadas, customizados para um destinatário específico, com aparência de credibilidade, visando a obter informações sensíveis (senhas, por exemplo) ou instalar *malwares*, que são programas maliciosos com efeitos gravosos sobre os sistemas afetados.

33 *Doxxing* significa a subtração maliciosa de informações acerca de alguma pessoa, seja em arquivos públicos ou hackeando computadores, com o propósito de assediar, intimidar ou extorquir, entre outros.

5. Discriminação algorítmica

Os algoritmos são treinados sobre os dados existentes, que, a seu turno, expressam comportamentos humanos passados e presentes, repletos de vieses e preconceitos, profundamente determinados por circunstâncias históricas, culturais e sociais (Prado, 2023, p. 162; Horta, 2019, p. 85-122). Tendem, por tal razão, a reproduzir estruturas sociais atuais e pretéritas de inclusão e exclusão. Nessa medida, dados sobre empregabilidade retratam uma menor contratação de mulheres, negros e indígenas, inclinação esta desprovida de relação com sua capacidade e produtividade, mas que pode induzir à reprodução de comportamentos futuros (Dastin, 2018)[34]; dados sobre segurança pública registram maior propensão à reincidência e violência envolvendo pessoas negras, não necessariamente porque sejam mais violentos, mas eventualmente porque vivem em contextos sociais mais adversos (Larson, 2016); dados sobre custos com a saúde tendem a superdimensionar os gastos de alguns grupos e minimizar os gastos de outros, por motivos não necessariamente relacionados às suas condições físicas[35]; dados sobre risco de crédito majorarão os riscos e, consequentemente, os custos de financiamento daqueles com menor *status* econômico e social, mesmo quando tenham logrado aprimorar suas condições, a depender das circunstâncias de coleta dos dados (Pasquale, 2016). Nessa medida, constata-se que alguns algoritmos de contratação podem tender a descartar mulheres, criminalizar homens negros e dificultar o acesso dos mais pobres ao crédito. Em tais condições, o modo de funcionar da IA pode ser profundamente reforçador de desigualdades existentes, em detrimento dos grupos mais vulneráveis da sociedade (Huq, 2020, p. 29-34; Silberg; Manyika, 2019, p. 3).

34 De fato, a Amazon deixou de utilizar um sistema de seleção de candidatos a emprego depois de identificar que tal sistema discriminava em desfavor da contratação de mulheres. A empresa constatou que a discriminação decorria do fato de que o sistema de IA fora treinado sobre um conjunto de dados sobre contratação recolhido ao longo dos últimos 10 anos, quando as mulheres estavam menos inseridas no mercado de trabalho. O sistema interpretou a menor presença de mulheres como se a contratação de homens fosse preferível, descartando candidatas mulheres.

35 No caso, os dados utilizados para treinar o algoritmo eram incompletos. Utilizaram-se dados sobre custos com pacientes brancos e negros para se estimar o alcance das suas necessidades de saúde. Os recursos empregados em pacientes negros eram inferiores àqueles empregados em pacientes brancos, não porque suas necessidades eram menores, mas porque tinham maior dificuldade de acesso ao serviço. Em razão disso, as necessidades de pacientes negros foram erradamente subdimensionadas pela IA (Obermeyer, 2019).

6. Questões sobre propriedade intelectual e direitos autorais

O modelo de negócio da IA suscita questões importantes acerca de direitos autorais e de propriedade intelectual. A quem pertencem os direitos autorais do amplo universo de canções, filmes, reportagens e conteúdos recolhidos pelas *big techs* com o propósito de alimentar suas IAs? A seus autores e criadores ou àqueles que passaram a empregá-los e explorá-los por meio de algoritmos? A IA generativa é alimentada com uma incrível quantidade de dados. No entanto, as respostas às indagações que lhe são formuladas vêm sem identificação da fonte e do autor. As discussões sobre esse tema vêm se acirrando e chegaram aos tribunais. Tome-se o exemplo da imprensa. Os conteúdos produzidos por empresas jornalísticas são recolhidos pelas empresas de IA, que os utiliza para treinar aplicativos que concorrem com os próprios veículos de imprensa na produção da informação[36]. A questão é objeto de uma ação judicial proposta pelo jornal *The New York Times* em face da OpenAI e da Microsoft (Grynbaum; Mac, 2023). Demanda semelhante envolve a Getty Images, empresa de mídia visual e fornecedora de imagens, e a Stability AI, empresa de inteligência artificial (Vincent, 2023)[37].

Não há como ser exaustivo na exploração dos riscos envolvidos no desenvolvimento da IA, pois são incontáveis as possibilidades a serem consideradas, fora aquelas que não somos sequer capazes de imaginar e antecipar. Mas há uma última preocupação que merece uma reflexão especial. Diz respeito ao que se denomina de "singularidade", termo empregado para identificar o risco de os computadores ganharem consciência, adquirirem vontade própria e se tornarem dominantes sobre a condição humana. Isso porque, sendo capazes de processar volume muito maior de dados em velocidade igualmente muito maior, se tiverem consciência e vontade se tornarão superiores a todos nós. O temor advém do fato de que os sistemas de IA podem se autoaperfeiçoar, atingindo a "superinteligência", dominando conhecimentos científicos, cultura geral e habilidades sociais que os colocariam acima dos melhores cérebros humanos.

Alguém cético das potencialidades humanas poderia até mesmo supor que uma superinteligência extra-humana teria maior capacidade de equacionar

36 Sobre a crise do modelo de negócio da imprensa e o impacto que produz na democracia, veja-se em: Minow, 2021, p. 35; Jackson, 2022 p. 280 *et seqs*; Barroso; Barroso, 2023; Campos Mello; Rudolf *apud* Cunha França; Casimiro, 2023.

37 A Getty Images alega que a Stability utilizou as imagens por ela produzidas para treinamento de um sistema de IA gerador de imagens denominado Stable Diffusion, sem autorização, violando seus direitos de propriedade intelectual e os direitos autorais de seus colaboradores, com o propósito de oferecer serviços semelhantes aos seus.

algumas das grandes questões não resolvidas da humanidade, como pobreza, desigualdade ou degradação ambiental. Mas nunca se poderia saber se essa inteligência fora de controle serviria à causa e aos valores da humanidade. Por isso mesmo, a governança da IA, doméstica e internacional, precisa estabelecer protocolos de segurança e parâmetros éticos destinados a administrar e mitigar esse risco. Se a tecnologia puder chegar a esse ponto – o que é colocado em dúvida por muitos cientistas – estará em jogo o próprio futuro da civilização e da humanidade.

Yuval Noah Harari (2023) faz um curioso comentário a respeito do tema. Segundo ele, em 2022, cerca de 700 dos mais relevantes cientistas e pesquisadores de IA foram indagados sobre os perigos dessa tecnologia impactar a própria existência humana ou provocar um expressivo desempoderamento. Metade deles respondeu que o risco seria de 10% ou mais. Diante disso, faz ele a pergunta fatídica: você entraria num avião se os engenheiros que o construíram dissessem que há um risco de 10% de ele cair? Se for isso mesmo, não dá para dormir tranquilo.

PARTE III
ALGUNS PRINCÍPIOS PARA REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL

1. Complexidades da regulação

Por tudo o que foi exposto até aqui, constata-se que a regulação da inteligência artificial se tornou imprescindível. Nada obstante, a tarefa não é singela e enfrenta desafios e complexidades. A seguir, procuramos identificar alguns deles.

A regulação precisa ser feita com o trem em movimento. Em março de 2023, mais de mil cientistas, pesquisadores e empreendedores assinaram uma carta aberta pedindo uma pausa no desenvolvimento dos sistemas mais avançados de IA, diante dos "profundos riscos para a sociedade e para a humanidade" que representavam. A pausa, por pelo menos seis meses, se destinaria a introduzir "um conjunto de protocolos de segurança compartilhados" (Future for Life Institute, 2023). As preocupações se justificavam plenamente, mas a suspensão das pesquisas não aconteceu. O trem continuou em alta velocidade. Até porque os avanços nessa área se tornaram objeto de disputa entre nações, pesquisadores e empreendedores. A carta, porém, reforçou as demandas por

governança, regulação, monitoramento e atenção para os impactos sociais, econômicos e políticos das novas tecnologias.

A velocidade das transformações é estonteante. Tal fato dificulta, imensamente, a previsibilidade do que está por vir e a apreensão das novas realidades em normas jurídicas, que correm o risco de se tornar obsoletas em pouco tempo. Não é difícil ilustrar o ponto. O telefone fixo tradicional levou 75 anos para atingir 100 milhões de usuários. O telefone móvel levou 16 anos. A Internet, 7 anos. Pois bem: o ChatGPT atingiu 100 milhões de usuários em dois meses (The Feed, 2023). Não é fácil para a legislação e a regulação acompanharem o ritmo das inovações.

Riscos da regulação excessiva. A regulação se tornou imprescindível, como assinalado acima, mas ela própria envolve riscos. Dois deles merecem destaque. O primeiro é o de que as restrições e a responsabilização civil não podem ser tão gravosas a ponto de inibir o ímpeto da inovação. Em segundo lugar, uma regulação desproporcional pode criar uma reserva de mercado para as empresas já estabelecidas, criando um fosso entre elas e a concorrência, agravando a concentração econômica nos grandes *players*. O conhecimento convencional vigente é que a regulação deve ter por foco os resultados, e não a pesquisa em si.

Assimetria de informação e de poder entre empresas e reguladores. A tecnologia da IA é controlada, sobretudo, pelas empresas envolvidas no seu desenvolvimento, que detêm conhecimento superior ao dos potenciais reguladores. A esse fato se soma que as empresas de tecnologia conhecidas como *big techs* são algumas das empresas mais valiosas do mundo, desfrutando de um poder econômico que é facilmente transformável em poder político. Tal poder ficou evidenciado quando da votação, no Congresso Nacional no Brasil, de projeto de lei que regulamentava a desinformação nas redes sociais. Algumas empresas de tecnologia deflagraram intensa campanha contra a medida, nas suas próprias plataformas e em *lobbies* no Congresso Nacional, conseguindo que o projeto fosse retirado de pauta (Rezende, 2024; Poder 360, 2023a; Poder 360, 2023b)[38].

Necessidade de harmonização global da regulação. A IA é uma tecnologia predominantemente privada, que não observa as fronteiras nacionais. As empresas operam globalmente e não costumam sequer ter sua sede nos principais centros de seus negócios. Dados podem ser coletados e alimentar o treinamento de sistemas em diferentes partes do mundo. Em tais condições,

38 Estima-se que não há grande interesse em se conter notícias falsas ou na moderação de conteúdo. Quanto mais notícias falsas, maior o engajamento do usuário e maior a interação nas redes, portanto, maior é a produção de dados, matéria-prima das *big techs*.

o modo de funcionar da IA coloca em xeque alguns elementos essenciais do direito, tal como o praticamos. Tais elementos são: a oponibilidade de direitos fundamentais e humanos aos Estados (e não propriamente a agentes privados) e o alcance das jurisdições nacionais, que encontram limite nas soberanias dos demais países. Além disso, o tratamento regulatório heterogêneo do tema, nos distintos países, pode gerar fuga de investimentos e obstáculos ao desenvolvimento tecnológico em Estados restritivos e representar um convite a uma ampla violação a direitos em locais mais permissivos.

2. Alguns esforços de regulação

No plano internacional, algumas iniciativas envolvendo proposições não vinculantes (*soft law*) foram marcantes. Entre elas, destacam-se: a) a Recomendação do Conselho sobre Inteligência Artificial, da OCDE (Organização para a Cooperação e o Desenvolvimento Econômico), de 2019 (OCDE, 2019)[39]; e b) a Recomendação sobre Ética na Inteligência Artificial, da Unesco, de 2021 (Unesco, 2023)[40],[41]. Ambos os documentos procuram responder aos riscos já indicados acima, são convergentes e complementares e reúnem princípios bastante gerais sobre a IA, a serem detalhados pelas normas domésticas dos respectivos países.

Em âmbito nacional, os Estados Unidos da América editaram, no final de 2023, uma longa *Executive Order* (EO) sobre IA[42]. Trata-se de uma normativa ampla, que alcança múltiplas áreas de risco da tecnologia, por meio da qual o presidente dos EUA se dirigiu essencialmente às agências federais, conforme sua expertise, determinando que estabelecessem *standards* e medidas para testar, assegurar a segurança e a confiabilidade da tecnologia, evitar fraudes, impedir a discriminação algorítmica, a violação a direitos fundamentais dos cidadãos, de consumidores, competidores e estudantes. A EO previu, ainda, a identificação dos conteúdos produzidos por IA com marcas d'água. Estabeleceu a definição de boas práticas e a realização de estudos sobre os impactos

39 Essa recomendação foi igualmente adotada pelo G-20.

40 O documento elenca os seguintes 10 princípios: 1 – Proporcionalidade e não produção de dano; 2 – Segurança; 3 – Justiça e não discriminação; 4 – Sustentabilidade; 5 – Privacidade e proteção de dados; 6 – Supervisão e determinação humanas; 7 – Transparência e explicabilidade; 8 – Compreensão e educação; 9 – Responsabilização e controle; 10 – Pluralidade de participantes, governança adaptável e colaboração.

41 A Organização das Nações Unidas adotou, ainda, Princípios para o Uso Ético da IA no Sistema das Nações Unidas, bastante similares àqueles objetos da recomendação da Unesco (Ceb, 2022).

42 A *Executive Order* é um diploma normativo, uma espécie de diretiva, editado pelo Presidente dos Estados Unidos da América, voltada à gestão do governo federal. É semelhante a um decreto, no Brasil. Encontra limites quanto à sua possibilidade de normatização, por não se tratar de uma lei produzida pelo Legislativo, e pode ser alterada por decisão do próximo presidente.

da IA nas relações de trabalho, com medidas para mitigá-los. Contemplou o financiamento para pesquisa e apoio a pequenas empresas no acesso a assistência técnica, a recursos e a mercado em IA, bem como a atração de novos talentos por meio de medidas de imigração. Determinou que desenvolvedores de modelos fundacionais que possam apresentar riscos para a segurança nacional, para a economia nacional e para a saúde pública notifiquem o Poder Púbico quando do treinamento dos seus sistemas e compartilhem com ele o resultado dos seus testes de segurança (*red-team safety tests*). E apelou ao Congresso para que aprovasse uma lei tutelando o direito à privacidade e protegendo os dados dos cidadãos.

Já a União Europeia (EU) aprovou, em março de 2024, o Ato da Inteligência Artificial (EU AI Act). A regulação proposta no âmbito da EU, diferentemente do que ocorre com a *Executive Order* norte-americana, caracteriza-se por estabelecer diretamente regras e sanções em matéria de desenvolvimento, implementação e operação da IA. Ela prevê, ainda, a atuação concentrada de determinados órgãos na sua vigilância e implementação. Tais normas são, contudo, proporcionais ao risco oferecido pela tecnologia (*risk based approach*) para pessoas e bens (União Europeia, 2021, 2023). Nessa linha, os sistemas são classificados em três níveis: a) sistemas sujeitos a riscos inaceitáveis, cuja implementação é proibida[43]; b) sistemas de alto risco, cuja implementação é permitida, desde que atendam a normas obrigatórias[44]; e c) sistemas de IA que não oferecem alto risco, para os quais se preveem incentivos à adoção voluntária de códigos de conduta, uma espécie de autorregulação (União Europeia, 2024).

No Brasil, tramitam no Congresso Nacional o Projeto de Lei (PL) n. 21/2020 e o PL n. 2.338/2023, havendo por aqui uma tendência de aproximação aos *standards* previstos nas propostas de norma da União Europeia (Brasil, 2023). Em linhas gerais, as propostas buscam: a) garantir direitos às pessoas diretamente afetadas pelos sistemas de IA; b) estabelecer responsabilidades de acordo com os níveis de riscos impostos por sistemas e algoritmos orientados por esse tipo de tecnologia; e c) estabelecer medidas de governança aplicáveis a empresas e a organizações que explorem esse campo.

43 Nessa categoria se inserem tecnologias de ranqueamento social (*social scoring*), identificação biométrica em locais públicos para fins de aplicação da lei (salvo exceções específicas), bem como práticas subliminares de manipulação de pessoas e/ou de exploração de vulnerabilidades de grupos vulneráveis.

44 Nessa categoria se incluem as tecnologias de ranqueamento social (*social scoring*), identificação biométrica em locais públicos para fins de aplicação da lei (salvo exceções específicas), bem como práticas subliminares de manipulação de pessoas e/ou de exploração de vulnerabilidades de grupos vulneráveis.

3. Algumas diretrizes

À luz de tudo o que foi exposto até aqui, é possível extrair alguns valores, princípios e objetivos que devem pautar a regulação da IA, para que essas tecnologias sirvam à causa da humanidade, potencializando-lhes os benefícios e minimizando os riscos. Tal regulação deve voltar-se à defesa dos direitos fundamentais, à proteção da democracia e à promoção da boa governança. A seguir, alguns elementos e aspectos ligados a cada uma dessas finalidades.

3.1 Defesa dos direitos fundamentais

a) Privacidade. O uso da IA deve respeitar os dados individuais das pessoas físicas e jurídicas, sem poder utilizá-los sem consentimento. A vigilância invasiva (*invasive surveillance*), como reconhecimento facial, biometria e monitoramento de localização deve ter emprego restrito e controlado. E, tendo em vista a vastidão de dados utilizados para se alimentar a IA, deve haver mecanismos adequados de segurança contra vazamentos.

b) Igualdade (não discriminação). A igualdade de todas as pessoas, em sua dimensão formal, material e de reconhecimento, é um dos mais valiosos pilares da civilização contemporânea. Já se alertou aqui, anteriormente, para os perigos da discriminação algorítmica. É preciso que a regulação da IA impeça que as pessoas sejam desequiparadas com base em categorias suspeitas, que exacerbem vulnerabilidades, como gênero, raça, orientação sexual, religião, idade e outras características. Há maus antecedentes nessa matéria (Dastin, 2018; Larson *et al*., 2016; Obermeyer *et al*., 2019; Heaven, 2021).

c) Liberdades. No que toca à autonomia individual, o uso da neurociência e da publicidade dirigida (*microtargeting*) tem o poder de manipular o comportamento e a vontade das pessoas, pelo sentimento do medo, do preconceito, da euforia e de outros vieses cognitivos, induzindo-as a comprar bens, contratar serviços ou adotar comportamentos contrários ao seu interesse, violando sua liberdade cognitiva ou autodeterminação mental. Além disso, o direito à informação, ao pluralismo de ideias e à liberdade de expressão podem ser comprometidos por algoritmos de recomendação ou de moderação, que filtram, direcionam e excluem conteúdos, em condutas equivalentes a uma censura privada.

3.2 Proteção da democracia

a) Combate à desinformação. A democracia é um regime de autogoverno coletivo, que pressupõe a participação esclarecida e bem informada dos cidadãos. Por isso mesmo, a circulação da desinformação e das teorias conspiratórias enganam ou geram medos infundados nas pessoas, comprometendo seu discernimento e suas escolhas. Como já observado, tudo isso é agravado pelas *deep fakes*, que simulam vídeos e falas inexistentes, com aparência de realidade. Todos nós somos educados para acreditar no que vemos e ouvimos. Manipulações dessa natureza quebram os paradigmas da experiência (Filimowicz *apud* Filimowicz, 2022, p. x e xi)[45] e são destrutivas da democracia.

b) Combate aos discursos de ódio. Desde que consagrado historicamente o sufrágio universal, a democracia envolve a participação igualitária de todas as pessoas. Discursos de ódio consistem em ataques a grupos vulneráveis, manifestações racistas, discriminatórias ou capacitistas relativamente a negros, *gays*, pessoas com deficiência e indígenas, entre outros. Ao pretender desqualificar, enfraquecer ou calar alguns grupos sociais, os discursos de ódio minam a proteção da dignidade humana e fragilizam a democracia.

c) Combate aos ataques às instituições democráticas. As redes sociais, auxiliadas pela IA, têm sido instrumentais na articulação de ataques às instituições democráticas, visando à sua desestabilização. Atos insurrecionais como o 6 de janeiro de 2021, nos Estados Unidos, ou o 8 de janeiro, no Brasil, com tentativas golpistas de desrespeito ao resultado das eleições colocam em risco a democracia e não podem ser tolerados (Ramonet, 2022).

3.3 Promoção da boa governança

À luz das recomendações e dos atos normativos internacionais, regionais e domésticos já referidos, e do amplo debate público em curso, na academia, na sociedade civil e na imprensa, é possível extrair alguns consensos sobrepostos no tocante à governança da IA, alinhavados nas cinco diretrizes expostas a seguir.

a) Centralidade do bem comum. A IA deve ser desenvolvida e estar orientada ao bem-estar das pessoas, dos países e do planeta. Seus benefícios devem ser

45 A experiência acumulada nos induz a compreender vídeos e registros fotográficos como elementos de atestação da veracidade das informações. Contudo, a IA baseada em tecnologias de aprendizado profundo permite a produção de mídias fraudulentas, mas ultrarrealistas, que confundem as mentes e permitem a produção sintética de falsas representações da realidade.

distribuídos de maneira justa entre todos e seus impactos negativos devem ser mitigados por meio da legislação e da regulação[46].

b) Governança plural. A governança da IA deve contemplar, em suas distintas etapas, com a proporcionalidade própria, a participação de um conjunto variado de atores, que inclui o Poder Público, cientistas e pesquisadores, sociedade civil, academia, empresas e entidades de direitos humanos. A diversidade de perspectivas e o sopesamento de valores e de interesses são muito importantes para a legitimidade das decisões e normatizações adequadas.

c) Transparência e explicabilidade. A transparência identifica o conhecimento mínimo do usuário sobre o funcionamento do sistema e a informação de que está interagindo com um sistema de IA. A explicabilidade significa tornar inteligíveis as razões das decisões tomadas, inclusive para permitir eventuais questionamentos dos resultados. Ambas as exigências se conjugam para mitigar preocupações com a precisão e a imparcialidade dos algoritmos, assim como para incentivar o uso responsável das tecnologias de automação (Bender, 2022, p. 12).

d) Segurança. Os sistemas de IA devem ser internamente seguros no sentido de evitar erros que produzam resultados indesejados, bem como devem, igualmente, estar protegidos contra ataques externos. A segurança no uso da IA inclui análise de impacto, cuidados com a qualidade dos dados, com a cybersegurança e o mapeamento dos processos e decisões que integram o ciclo de vida da IA (*traceability*).

e) Controle e responsabilidade. A supervisão ou controle humanos são fundamentais para que a IA esteja operando dentro das balizas da legalidade, da ética e da justiça. Apesar da relativa autonomia nos seus processos decisórios, a responsabilidade será sempre de uma pessoa física ou jurídica. Em caso de uso indevido ou malicioso, uma delas ou ambas estarão sujeitas à responsabilização civil, administrativa e penal.

46 A regulação deve ser vista como uma condição necessária, mas insuficiente. O tratamento dos riscos associados à inteligência artificial, nesse sentido, perpassa a dimensão do direito para alcançar, igualmente, outros campos, entre os quais se destaca a ética aplicada à economia e à programação. Para Lucrecio Rebollo: "Conceber o direito como única forma de ordenar e equalizar a sociedade digital é um erro grave. O direito deve ser, como sempre foi, uma forma de se resolverem os conflitos sociais com uma perspectiva de bem comum, mas em todos os casos ele necessita da colaboração de outras áreas do conhecimento, de todos os elementos que conformam a estrutura social" (Rebollo Delgado, 2023, p. 52).

Conclusão

O papel do conhecimento é confortar os aflitos e afligir os confortados (Shedden, 2014)[47]. O presente artigo tem a pretensão de haver cumprido esse papel. A inteligência artificial, como aqui demonstrado, apresenta potencialidades e riscos em quase todas as áreas em que pode ser aplicada. No plano político, pode ajudar a aprimorar o sistema representativo e a captar melhor o sentimento e a vontade dos cidadãos. Mas pode, também, massificar a desinformação, os discursos de ódio e as teorias conspiratórias, enganando os eleitores, fragilizando grupos vulneráveis ou disseminando temores infundados, extraindo o pior das pessoas.

No plano econômico, a IA pode contribuir para o aumento da produtividade em áreas diversas, do agronegócio à indústria, bem como aprimorar significativamente o setor de serviços. Mas pode, também, concentrar riquezas nos setores mais favorecidos e nas nações mais ricas, aumentando a desigualdade no mundo. No plano social, pode ser um instrumento importante no equacionamento de problemas ligados à pobreza e às desigualdades injustas, mas pode, por outro lado, levar ao desemprego massas de trabalhadores. Existem, também, dualidades éticas. Uma maior compreensão da natureza humana pode elevar o patamar humanístico ou espiritual no mundo, mas não é descartável a perda da centralidade da pessoa humana.

Em suma, vivemos uma era de ambiguidades e de escolhas decisivas. Na visão dos autores, a história do mundo tem sido um fluxo constante – embora não linear – na direção do bem, da justiça e do avanço civilizatório. Viemos de tempos de asperezas, sacrifícios humanos e despotismos, até chegarmos à era dos direitos humanos. Por essa razão, é possível termos uma visão e uma atitude construtivas em relação à inteligência artificial. Sem medos paralisantes, mas, também, sem ingenuidades ou fantasias. Vamos precisar de legislação, de regulação e, sobretudo, de educação e conscientização de cientistas, empresas e cidadãos para não nos perder pelo caminho. E, como já referido, a bússola, o rumo indicado pelas estrelas, são os valores que conduzem à vida boa: virtude, razão prática e coragem moral. Se perdermos as referências do bem, da justiça e da dignidade humana, aí seria o caso, mesmo, de deixar as máquinas tomarem conta e apostar que poderão fazer melhor.

47 Esta frase é uma paráfrase de *The job of the newspaper is to comfort the afflicted and afflict the comfortable*, atribuída a um personagem de ficção – Mr. Dooley –, criado pelo jornalista Finley Peter Dunne, do Chicago Evening Post.

Mas não há de ser assim. Talvez, paradoxalmente, a inteligência artificial possa ajudar a resgatar e a aprofundar a nossa própria humanidade, valorizando a empatia, a fraternidade, a solidariedade, a alegria, a capacidade de amar e demais atributos que sempre nos diferenciarão de máquinas.

Referências

ACEMOGLU, Daron e SIMON, Johnson. *Power and progress*. N. York, Public Affairs, 2023.

ALVIM, Frederico Franco; ZILIO, Rodrigo López; CARVALHO, Volgane Oliveira. *Guerras cognitivas na arena eleitoral*: o controle judicial da desinformação. Rio de Janeiro: Lumen Juris, 2023, p. 69.

ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren. Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks. *ProPublica*, 23 maio 2016.

ARIEL GUSTAVO. Are artificial intelligence courts a discrimination risk? *European AI alliance*, 31 ago. 2021. Disponível em: https://futurium.ec.europa.eu/en/european-ai-alliance/open-discussion/are-artificial-intelligence-courts-discrimination-risk. Acesso em: 3 abr. 2014.

ARIYARATNE, Hasala. *ChatGPT and intermediar liability*: why section 230 does not and should not protect generative algorithms, p. 4-5, 16 maio 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4422583.

BARROSO, Luís Roberto. Revolução tecnológica, crise da democracia e mudança climática. *Revista Estudos Institucionais*, v. 5, n. 3, p. 1262, 2019.

BARROSO, Luís Roberto; BARROSO, Luna van Brussel. Democracia, mídias sociais e liberdade de expressão: ódio, mentiras e a busca pela verdade possível. *Direitos fundamentais e Justiça*, ano 17, n. 49, p. 285-311, jul./dez. 2023.

BARROSO, Luís Roberto; BARROSO, Luna van Brussel. Democracy, Social Media, and Freedom of Expression: Hate, Lies, and the Search for the Possible Truth, Chicago *Journal of International Law*, v. 24, n. 50, 2023.

BECKETT, Lois; PAUL, Kari. Bargaining for our very existence: why the battle over AI is being fought in Hollywood. *The Guardian*, 22 jul. 2023. Acesso em: 17 jan. 2024.

BEIGUELMAN, Giselle. *Políticas da imagem*: vigilância e resistência na dadosfera. São Paulo: Ubu, 2021.

BENDER, Sarah M. L. Algorithmic Elections. *Michigan Law Review*, v. 121, n. 3, p. 489-522, 2022.

BERGHEL, Hal. Malice Domestic: The Cambridge Analytica Dystopia. *Computer*, p. 84-89, maio 2018.

BOMMASANI, Richi *et al*. On the Opportunities and Tasks of Foundation Models. Center for Research on Foundation Models. *Stanford University*, p. 4-12, 12 jul. 2022.

BRASIL. Câmara dos Deputados. *PL 21/2020*, iniciativa Deputado Eduardo Bismark, situação: aprovado com alterações no Plenário e remetido ao Senado. Disponível em: https://www.camara.leg.br/propostas-legislativas/2236340. Acesso em: 17 fev. 2024.

BRASIL. Senado Federal. *PL 2338/2023*, inciativa Senador Rodrigo Pacheco, relator atual Senador Eduardo Gomes, situação: com a relatoria. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233. Acesso em: 17 fev. 2024.

BROWN, Sara. Machine learning, explained. *MIT Management Sloan School*. 21 abr. 2021. Disponível em: https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained?utm_source=mitsloangooglep&utm_medium=social&utm_campaign=machinelearnexp&gad_source=1&gclid=Cj0KCQiAtaOtBhCwARIsAN_x-3KnfPNYty2tnOgUTP0F_NMirqdswn7etv0WLC6YxWMNvm3jH1sxEJwaAp0REALw_wcB. Acesso em: 18 jan. 2024.

CACZAN, Luciana. Tetraplégico recupera movimentos após implantar chips de inteligência artificial no cérebro. *CNN Brasil*, 20 jul. 2023. Disponível em: https://www.cnnbrasil.com.br/tecnologia/tetraplegico-recupera-movimentos-apos-implantar-chips-de-inteligencia-artificial-no-cerebro/. Acesso em: 14 fev. 2024.

CAMPOS MELLO, Patrícia. *A máquina do ódio*. Companhia das Letras, 2020.

CHATGPT. *Página Inicial*, [20--]. Disponível em: https://chat.openai.com/auth/login. Acesso em: 17 jan. 2024.

CITRON, Danielle; SOLOVE, Daniel J. Privacy Harms. *Boston University Law Review*, v. 102, p. 793-863, 2022.

COECKELBERGH, Mark. *Ética na inteligência artificial*. Rio de Janeiro: Ubu, 2023.

COX, Joseph. Data Broker Is Selling Location Data of People Who Visit Abortion Clinics. *Vice*, 2022. Disponível em: https://www.vice.com/en/article/m7vzjb/location-data-abortion-clinicssafegraph-planned-parenthood. Acesso em: 17 jan. 2024.

DASTIN, Jeffrey. Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women, *Reuters*, 10 out. 2018. Disponível em: https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G/. Acesso em: 12 fev. 2024.

DAVENPORT, Thomas; KALACOTA, Ravi, The potential for Artificial Intelligence in healthcare. *Future Healthcare Journal*. v. 6, n. 94, 2019.

DEGLI-ESPOSTI, Sara. *La ética de la inteligencia artificial*. Madrid: Catarata, 2023.

DILSIZIAN, Steven E.; SIEGEL, Eliot L. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. *Current cardiology reports*, v. 16, p. 1-8, 2014. Disponível em: https://doi.org/10.1007/s11886-013-0441-8. Acesso em: 14 fev. 2024.

DREYFUS, Hubert L.; DREYFUS, Stuart E. Making a Mind Versus Modeling the Brain: Artificial Intelligence Back at a Branchpoint. *Dædalus*, v. 1, n. 117, p. 15–44, 1988. Disponível em: https://www.amacad.org/sites/default/files/daedalus/downloads/Daedalus_Wi98_Artificial-Intelligence.pdf. Acesso em: 17 jan. 2024.

DURÃES, Uesley. Reconhecimento facial: erros expõem falta de transparência e viés racista. *Uol*, 28 abr. 2024. Disponível em: https://noticias.uol.com.br/cotidiano/ultimas-noticias/2024/04/28/reconhecimento-facial-erros-falta-de-transparencia.htm. Acesso em: 7 maio 2024.

ESTADOS UNIDOS DA AMÉRICA. The New York Times v. Microsoft Corporation, OpenAI Inc., OpenAI LP, OpenAI GP LLC, OpenAI LLC, OpenAI OpCo LLC, OpenAI Global LLC, OAI Corporation, LLC, OpenAI Holdings, LLC. Federal District Court of Manhattan, 27 dez. 2023. Disponível em: https://nytco-assets.nytimes.com/2023/12/NYT_Complaint_Dec2023.pdf. Acesso em: 12 fev. 2024.

EUBANKS, Virgínia. The Allegheny Algorithm. *In*: EUBANKS, Virgínia. *Automating Inequality*: *How High-Tech Tools Profile, Police, and Punish the Poor*. Nova Iorque: St. Martin's Press, 2015. Disponível em: https://acrobat.adobe.com/link/review?uri=urn:aaid:scds:US:92531b35-dd7b-3c72-8721-ee7781ddb786. Acesso em: 17 jan. 2024.

FILIMOWICZ, Michael. Introduction. *In*: FILIMOWICZ, Michael. *Deep Fakes*. Algorithms and Society. New York: Routeledge, 2022, p. X-XI.

FISHER, Max. *A máquina do caos*. Como as redes sociais reprogramaram nossa mente e nosso mundo. São Paulo: Todavia, 2023.

FREY, Carl B.; OSBOURNE, Michael A. The Future of Employment: How Susceptible Are Jobs to Computerisation?, *Future of Humanity Institute*, Jan. 2017. Disponível em: accessible through Harvard at https://www-sciencedirect-com.ezp-prod1.hul.harvard.edu/science/article/pii/S0040162516302244. Acesso em: 11 fev. 2024.

FUTURE OF LIFE. *Pause AI Giant experiments*: an open letter, 22 mar. 2023. Disponível em: https://futureoflife.org/open-letter/pause-giant-ai-experiments/. Acesso em: 12 fev. 2024.

GALAZ, Victor *et al*. Artificial intelligence, systemic risks, and Sustainability. *Technology in Society*, v. 67, p. 1-10, set. 2021.

GARCÍA-MATÍN, E. *et al*. Estimation of energy consumption in machine learning, J. Parallel Distributed. *Computing*, v. 134, p. 75-88, 2019.

GRYNBAUM, Michael M.; MAC, Ryan. The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work. *The New York Times*, 27 dez. 2023. Disponível em: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html. Acesso em: 12 fev. 2024.

GUADIÁN, Carlos. Cómo va a afectar la Inteligencia Artificial las elecciones en 2024. *CludPad*, 25 de janeiro de 2024. Disponível em: [https://carlosguadian.substack.com/p/como-va-a-afectar-la-inteligencia?utm_source=post-email-title&publication_id=259698&post_id=141029307&utm_campaign=email-post-title&isFreemail=true&r=4n36v&utm_medium=email]. Acesso em: 21 abr. 2024.

HACKER, Philipp; ENGEL, Andreas; MAUER, Marco. Regulating ChatGPT and Other Large Generative Models. Working Paper**.** *ACM Conference on Fairness, Accountability, and Transparency*, 12 maio 2023. Disponível em: https://doi.org/10.1145/3593013.3594067. Acesso em: 18 jan. 2023.

HAO, Karen. What is machine learning? *MIT Technology Review*, 17 nov. 2018. Disponível em: https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/. Acesso em: 18 jan. 2024.

HARARI, Yuval Noah. You can have the blue pill or the red pill, and we are out of blue pills. *New York Times*, 24 mar. 2023. Disponível em: https://www.nytimes.com/2023/03/24/opinion/yuval-harari-ai-chatgpt.html. Acesso em: 28 maio 2024.

HARARI, Yuval. Yuval Noah Harari argues that AI has hacked the operating system of human civilisation. *The Economist*, 28 abr. 2023. Disponível em: https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation?dclid=CP691aS5kYMDFaWNrAId4O0GXA&utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=19495686130&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gad_source=1&gclid=CjwKCAjwupGyBhBBEiwA0UcqaHf0aSx4CKaU2YdHw_bw_3ep3pJdU9e8T4cykjwu-Y6E9eI6yr7OPxoCvEUQAvD_BwE&gclsrc=aw.ds. Acesso: 31 mar. 2024.

HARMS, William. Meeting online leads to happier, more enduring marriages. *UChicago News*, 3 jun. 2013.

HEAVEN, Will Douglas. Predictive Policing is Still Racist—Whatever Data it Uses, *MIT Technology Review*, 5 fev. 2021, https://www.technologyreview.com/2021/02/05/1017560/predictive-policing-racist-algorithmic-bias-data-crime-predpol. Acesso em: 7 maio 2024.

HEAWOOD, Jonathan. Pseudo-public political speech: Democratic implications of the Cambridge Analytica scandal**.** *Information Polity*, v. 23, p. 429-434, 2018.

HERE.AFTER. *Your stories and voice*: forever, [20--]. Página inicial. Disponível em: https://www.hereafter.ai/. Acesso em: 17 jan. 2024.

HOROWITZ, Jason. Who Will Take Care of Italy's Older People? Robots, Maybe. *The New York Times*, 25 mar. 2023. Disponível em: https://www.nytimes.com/2023/03/25/world/europe/who-will-take-care-of-italys-older-people-robots-maybe.html. Acesso em: 17 jan. 2024.

HORTA, Ricardo Lins. Por que existem vieses cognitivos na tomada de decisão judicial? A contribuição da psicologia e das neurociências para o

debate jurídico. *Revista Brasileira de Políticas Públicas*, v. 9, n. 3, dez. 2019, p. 85-122.

HUQ, Aziz. Constitutional Rights in the Machine Learning State. *Cornell Law Review*, v. 105, 2020. Disponível em: https://ssrn.com/abstract=3613282. Acesso em: 11 fev. 2024.

HUTSON, Matthew. Can we stop runway A.I.? Technologists warn about the dangers of the so-called singularity. But can anything actually be done to prevent it? *The New Yorker*, 16 maio 2023. Disponível em: https://www.newyorker.com/science/annals-of-artificial-intelligence/can-we-stop-the-singularity. Acesso em: 17 fev. 2024.

INNER CIRCLE. *Você tá a 1 clique de conhecer o crush dos seus sonhos*. Página Inicial. Disponível em: https://m.theinnercircle.co/?utm_campaign=brbrand2&campaignid=14801829816&source=google&medium=cpc&gad_source=1&gclid=CjwKCAiAkp6tBhB5EiwANTCx1P65xV0LpMJQ_D8-3aqirg454qwI6ddMMISOgOY1f3C-eQjON337DBoCOicQAvD_BwE. Acesso em: 17 jan. 2024.

JACKSON, Vicki C. Knowledge Institutions in Constitutional Democracy: reflections on "the press", *Journal of Media Law*, v. 14, n. 2, 2022.

JONES, Elliot. Explainer: What is a foundation model? *Ada Lovelace Institute*, 17 jul. 2023. Disponível em: https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/#table1. Acesso em: 18 jan. 2024.

KAKUTANI, Michiko. *A morte da verdade*. Notas sobre a mentira na era Trump. Rio de Janeiro: Intrínseca, 2018.

KATYAL, Sonia K. Democracy and Distrust in an Era of Artificial Intelligence. *Daedalus*, n. 151, v. 2, p. 322–334, 2022.

KEYNES, John Maynard. *Economic Possibilities for our Grandchildren*, 1930. Disponível em: https://www.aspeninstitute.org/wp-content/uploads/files/content/upload/Intro_and_Section_I.pdf. Acesso em: 12 fev. 2024.

KISSINGER, Henry A.; SCHMIDT, Eric; HUTTENLOCHER, Daniel. *A era da IA*. Rio de Janeiro: Alta Books, 2023.

KISSINGER, Henry A.; SCHMIDT, Eric; HUTTENLOCHER, Daniel. *The age of AI and our human future*. Nova York: Little, Brown and Company, 2021.

KISSINGER, Henry, SCHMIDT, Eric Schmidt e HUTTENLOCHER, Daniel. ChatGPT Heralds an Intellectual Revolution. *Wall Street Journal*, 24 fev. 2023.

Disponível em: https://www.wsj.com/articles/chatgpt-heralds-an-intellectual-revolution-enlightenment-artificial-intelligence-homo-technicus-technology-cognition-morality-philosophy-774331c6. Acesso em: 21 maio 2024.

KLARE, Michael T. UN to address autonomous weapons systems. *Arms Control Association*, dez. 2023. Disponível em: https://www.armscontrol.org/act/2023-12/news/un-address-autonomous-weapons-systems. Acesso em: 28 mar. 2024.

LARSON, Jeff *et al*. How We Analyzed the Compas Recidivism Algorithm. *ProPublica*, 23 maio 2016. Disponível em: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Acesso em: 12 fev. 2024.

LEE, Kai-Fu. *Inteligência artificial*. Como os robôs estão mudando o mundo, a forma como amamos, nos relacionamentos, trabalhamos e vivemos. Rio de Janeiro: Globo Livros, 2019.

LENHARO, Mariana. AI consciousness: scientists say we urgently need answers. *Nature*, 21 dez. 2023.

LOHR, Steve. Generative A.I.'s Biggest Impact Will Be in Banking and Tech, Report Says. *The New York Times*, 1 fev. 2024. Disponível em: https://www.nytimes.com/2024/02/01/business/ai-impact-jobs.html?unlocked_article_code=1.SE0.-W1k.6YLx1J6GM3Uc&smid=wa-share. Acesso em: 11 fev. 2024.

MAHESHWARI, Sapha. A.I. Fuels a New Era of Product Placement. *The New York Times*. The New York Times, 1 fev. 2024. Disponível em: https://www.nytimes.com/2024/02/01/business/media/artificial-intelligence-product-placement.html?unlocked_article_code=1.SE0.p-JV.iFEjWlI2qW-4&smid=wa-share. Acesso em: 11 fev. 2024.

MAKHNOUMI, Ali. How AI Could Potentially Manipulate Consumers. *Duke Fuqua School of Business*, 10 jan. 2024. Disponível em: https://www.fuqua.duke.edu/duke-fuqua-insights/how-ai-could-potentially-manipulate-consumers. Acesso em: 17 fev. 2024.

MANYIKA, James. Getting AI right: Introductory notes on AI & society. *Daedalus,* v. 151, n. 2, p. 5-27, 2022.

MCCARTHY, John; MINSKY, Marvin L.; ROCHESTER, Nathaniel; SHANNON, Claude E. *A Proposal for the Dartmouth Summer Research Project on Artificial*

Intelligence. 31 ago. 1955. Disponível em: http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf. Acesso em: 17 jan. 2024.

MELLO, Patrícia Perrone Campos Mello; RUDOLF, Renata H. S. B. A. Redes Sociais e Democracia: Disrupção Tecnológica, Erosão Democrática e Novas Perspectivas. *In*: FRANÇA, Eduarda Peixoto da Cunha; CASIMIRO, Matheus (Org.). *Direito e Política*: Um Diálogo Possível? Londrina: THOTH Editora, p. 53-78, 2023.

MINDBANK.AI go beyond. *Duplicate yourself with a personal digital twin*, [20--]. Página inicial. Disponível em: https://www.mindbank.ai/. Acesso em: 17 jan. 2024.

MINOW, Martha. *Saving the Press*: Why the constitution calls for government action to preserve freedom of speech. Oxford University Press, 2021.

MOHAMMAD, Shabina; JAN, Raghad A.; ALSAEDI, Saba L. Symptons, Mechanisms, and Treatments of Video Game Addiction. *Cureus*, v. 15, n. 3, 31, mar. 2023. Disponível em: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10065366/#:~:text=Video%20game%20addiction%20is%20defined,in%20many%20aspects%20of%20life. Acesso em: 17 fev. 2024.

MOROZOV, Evgeny. *Big Tech*: *a ascensão dos dados e a morte da política*. São Paulo: Ubu, 2018.

MORRISON, Sara. This outed priest's story is a warning for everyone about the need for data privacy laws. *Vox*, 21 jul. 2021. Disponível em: https://www.vox.com/recode/22587248/grindr-app-location-data-outed-priest-jeffrey-burrill-pillar-data-harvesting. Acesso em: 17 fev. 2024.

MUÑOZ VELA, José Manuel. *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial*. Un enfoque de seguridad física, lógica, moral y jurídica. Pamplona: Arazandi, 2022.

NAÇÕES UNIDAS. Chief Executives Board for Coordination. *Principles for the Ethical Use of Artificial Intelligence in the United Nations System*, 20 set. 2022. Disponível em: https://unsceb.org/sites/default/files/2022-09/Principles%20for%20the%20Ethical%20Use%20of%20AI%20in%20the%20UN%20System_1.pdf. Acesso em: 15 maio 2024.

NUNES, Dierle José Coelho; ANDRADE, Otávio Morato de. O uso da inteligência artificial explicável enquanto ferramenta para compreender decisões automatizadas: possível caminho para aumentar a legitimidade e

confiabilidade de modelos algorítmicos? *Revista Eletrônica do Curso de Direito da UFSM*, v. 18, n. 1, 2023.

OBERMEYER, Ziad *et al*. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. *Science*, 25 out. 2019. Disponível em: https://www.science.org/doi/full/10.1126/science.aax2342. Acesso em: 12 fev. 2024.

OCDE. Recommendation of the Council on Artificial Intelligence. *OECD/LEGAL/0449, 2019*. Disponível em: https://oecd.ai/en/wonk/documents/g20-ai-principles. Acesso em: 13 fev. 2024.

OECD. *Opportunities, guidelines and guardrails for effective and equitable use of AI in education*. Paris: OECD Publishing, 2023. Disponível em:https://www.oecd.org/education/ceri/Opportunities,%20guidelines%20and%20guardrails%20for%20effective%20and%20equitable%20use%20of%20AI%20in%20education.pdf. Acesso em: 31 mar. 2024.

PASQUALE, Frank. *The Black BoX Society*: The Secr*et Al*gorithms That Control Money and Information. Cambridge: Harvard University Press, 2016. E-book.

PAUWELS, Eleonore. Artificial Intelligence and Data Capture Technologies in Violence and Conflict Prevention. *Global Center on Cooperative Security*: *Policy Brief*, set. 2020. Disponível em: https://www.globalcenter.org/wp-content/uploads/GCCS_AIData_PB_H-1.pdf. Acesso em: 17 fev. 2024.

PODER 360. *Google inclui texto contra PL da fake news na página inicial do buscador*. Big tech exibe link para artigo nomeado "PL das fake news pode aumentar a confusão entre o que é verdade ou mentira no Brasil", 1 maio 2023b. Disponível em: https://www.poder360.com.br/tecnologia/google-inclui-texto-contra-pl-das-fake-news-na-home-do-buscador/. Acesso em: 19 jan. 2024.

PODER 360. *PL das fake news dá "poderes de censura" ao governo*, diz Telegram, 9 maio 2023a. Disponível em: https://www.poder360.com.br/tecnologia/pl-das-fake-news-da-poderes-de-censura-ao-governo-diz-telegram/. Acesso em: 19 jan. 2024.

PORTO, Fábio Ribeiro; ARAÚJO, Walter Shuenquener; e GABRIEL, *Inteligência Artificial generativa no Direito*, 2024. (mimeo).

PRADO, Michelle. *Tempestade ideológica*. Bolsonarismo: a altright e o populismo iliberal no Brasil. São Paulo: Todos Livros, 2023.

RAMIREZ, Vanessa Bates. Grief tech uses AI to give you (and your loved ones) digital immorality. *Singularity hub*, 16 ago. 2023. Disponível em: https://singularityhub.com/2023/08/16/grief-tech-uses-ai-to-give-you-and-your-loved-ones-digital-immortality/.Acesso em: 17 jan. 2024.

RAMONET, Ignacio. *La era del conspiracionismo***:** Trump, el culto a la mentira y el asalto al Capitolio. Buenos Aires: Siglo XXI, 2022.

RE, Richard M.; SOLOW-NIEDERMAN, Alicia. Developing Artificially Intelligent Justice. *Stan. Tech. L. Rev.*, n. 22, p. 242-289, 2019.

REBOLLO DELGADO, Clicerio. *Inteligencia artificial y derechos fundamentales*. Madrid: Dykinson, 2023.

REINO UNIDO. Hight Copurt of Justice Business and Property Courts of England and Wales. *EWHC 30390*. Getty Images (US) Inc, Getty Images International UC, Letty Images UK Limited, Getty Images Devco UK Limited, Istockphoto LP, Thomas M Barwick Ink v. Stability AI Ltd. Case No. IL-2023-000007, 1 dez. 2023. Disponível em: https://www.documentcloud.org/documents/24183636-getty-images-v-stability-ai-uk-ruling. Acesso em: 12 fev. 2024.

REZENDE, Constança. PF conclui que Google e Telegram agiram de modo abusivo contra PL das Fake News. Empresas, que lançaram ofensiva contra projeto no ano passado, negam irregularidades. *Folha de São Paulo*, 31 jan. 2024. Disponível em: https://www1.folha.uol.com.br/poder/2024/01/pf-conclui-que-google-e-telegram-agiram-de-modo-abusivo-contra-pl-das-fake-news.shtml?utm_source=sharenativo&utm_medium=social&utm_campaign=sharenativo. Acesso em: 31 jan. 2024.

RUSSEL, Stuart; PERSET, Karine; MARKO, Grobelnik. Updates to the OECD's definition of an AI system explained. *OCDE AI*: policy observatory, 29 nov. 2023. Disponível em: https://oecd.ai/en/wonk/ai-system-definition-update. Acesso em: 17 jan. 2024.

RUSSELL, Stuart. *Inteligência artificial a nosso favor*. Como manter o controle sobre a tecnologia. São Paulo: Cia das Letras, 2021.

SCHICK, Nina. *Deep Fakes and the Infocalypse*. What You Urgently Need do Know.London: Monoray, 2020, p. 11.

SCHMIDT, Albrecht. Augmenting Human Intellect and Amplifying Perception and Cognition. *IEEE Pervasive Computing*, v. 16, n. 1, p. 6-10, jan./mar. 2017.

SHEDDEN, David. Today in Media History: Mr. Dooley: 'The job of the newspaper is to comfort the afflicted and afflict the comfortable'. *Poynter*, 7 out. 2014. Disponível em: https://www.poynter.org/reporting-editing/2014/today-in-media-history-mr-dooley-the-job-of-the-newspaper-is-to-comfort-the-afflicted-and-afflict-the-comfortable/. Acesso em: 7 maio 2024.

SILBERG, Jake; MANYIKA, James. Notes from the AI Frontier: Tackling Bias in AI (and in Humans). *McKinsey Global Institute*, jun. 2019. Disponível em: https://www.mckinsey.com/featuredinsights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans#. Acesso em: 17 jan. 2024.

SUNSTEIN, Cass R. Governing by algorithm? No noise and (potentially) less bias. *Duke Law Journal*, v. 71, n. 6, p. 1175–1206, 2022.

TAMBIAMA, Madiega. Briefing: EU Legislation in Progress. Artificial Intelligence Act. EPRA – European Parliamentary Research Service, jun. 2023.

TAULLI, Tom. *Introdução à Inteligência Artificial*. Uma abordagem não técnica. São Paulo: Novatec, 2020.

THE FEED. ChatGPT witnesses massive rise, chatbot gains 100 million users in two months. *Economic Times*, 5 mar. 2023. Disponível em: https://economictimes.indiatimes.com/news/new-updates/chatgpt-witnesses-massive-rise-chatbot-gains-100-million-users-in-two-months/articleshow/98428443.cms?from=mdr. Acesso: 31 mar. 2024.

TINDER. *Página Inicial*, [20--]. Disponível em: https://tinder.com/pt. Acesso em: 17 jan. 2024.

TROPIANO, Dolores. More internet marriages are leading to happy marriages. *Arizona State University News*, 20 jan. 2023. Disponível em: https://news.asu.edu/20230119-university-news-more-internet-matches-are-leading-happy-marriages. Acesso em: 31 mar. 2024.

TUFEKCI, Zeynep. We need to take our privacy back. *The New York Times*, 2022. Disponível em: https://www.nytimes.com/2022/05/19/opinion/privacy-technology-data.html. Acesso em: 17 jan. 2024.

UNESCO. *AI and education*: a guide for policy makers. 2021. Disponível em: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1196&id=p::usmarcdef_0000376709&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_761bcdad-d1e3-40c9-819d-03c4ac725f26%3F_%3D376709eng.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000376709/PDF/376709eng.pdf#AI%20in%20education_pages.indd%3A.14084%3A1005. Acesso em: 31 mar. 2024.

UNESCO. *Recommendation on the ethics of artificial intelligence*, 23 nov. 2021. Disponível em: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence. Acesso em: 13 fev. 2024.

UNESCO. *Recommendation on the ethics of artificial intelligence*, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000380455#:~:text=AI%20actors%20and%20Member%20States,law%2C%20in%20particular%20Member%20States'. Acesso em: 26 mar. 2024.

UNIÃO EUROPEIA. Comissão Europeia. Directorate-General for Research and Innovation. *Quarterly R&I literature review*: The impact of AI on R&I, n. Q2, 11 jul. 2023a.

UNIÃO EUROPEIA. Conselho da Europa. Artificial Intelligence and Electoral Integrity. *Concept Paper*, 2022. Disponível em: https://www.coe.int/en/web/electoral-management-bodies-conference/concept-paper-2022. Acesso em: 21 abr. 2024.

UNIÃO EUROPEIA. *EU Artificial Intelligence Act*, 2024. Disponível: https://artificialintelligenceact.eu/the-act/. Acesso em: 15 maio 2024. Acesso em: 14 fev. 2024.

UNIÃO EUROPEIA. European Commission. *Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules in Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union Legislative Acts*, 21 abr. 2021. Disponível: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206. Acesso em: 15 maio 2024.

UNIÃO EUROPEIA. Parlamento Europeu. *Artificial Intelligence Act*: deal on comprehensive rules for trustworthy AI, 12 set. 2023b. Disponível em: https://artificialintelligenceact.eu/the-act/. Acesso em: 14 fev. 2024.

URUEÑA, René. Regulating the Algorithmic Welfare State in Latin America. *Max Planck Institute for Comparative Public Law & International Law* (MPIL). Research Paper No. 2023-27, 20 dez. 2023.

UUK, Risto; GUTIERREZ, Carlos Ignacio; TAMKIN, Alex. Operationalising the Definition of General Purpose AI Systems: Assessing Four Approaches. Working Paper. *Stanford University*, 1 jun. 2023.

VERMA, Pranshu. The never-ending quest to predict crime using AI. *The Washington Post*, 15 jul. 2022.

VINCENT, Jame. Getty Images is suing the creators of AI art tool Stable Difusion for scraping its content. *The Verge*, 17 jan. 2023. Disponível em: https://www.theverge.com/2023/1/17/23558516/ai-art-copyright-stable-diffusion-getty-images-lawsuit. Acesso em: 12 fev. 2024.

VLACHOS, Scott. The link between mis-, dis-, and malinformation and domestic extremism. *Council for Emerging National Security Affairs*, June 2022. Disponível em: MDM_22.617b.pdf (censa.net). Acesso: 21 abr. 2024.

WEBB, Michael. *The Impact of Artificial Intelligence on the Labor Market*, 6 nov. 2019. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3482150. Acesso em: 11 fev. 2024.

WILLIAMS, Matthew. *A ciência do ódio*. Rio de Janeiro: Globo Livros, 2021, p. 207.

WINSTON, Patrick Henry. *Artificial intelligence desmystified*. Minuta de 30 set. 2018. (mimeo).

ZUBOFF, Shoshana. Surveillance Capitalism or Democracy? The Death Match of Institutional Orders and Politics of Knowledge in Our Information Civilization. *Organization Theory*, v. 3, p. 1-79, 2022.