IX ENCONTRO VIRTUAL DO CONPEDI

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS V

D598

Direito, governança e novas tecnologias V [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Vivianne Rigoldi; Jéssica Fachin; Rosane Teresinha Porto – Florianópolis: CONPEDI, 2026.

Inclui bibliografia
ISBN: 978-65-5274-589-7
Modo de acesso: www.conpedi.org.br em publicações
Tema: Constituição, Processo e Justiça Social: o papel da jurisdição constitucional na efetivação de direitos

2. Direito. 3. Governança e novas tecnologias. IX Encontro Virtual do CONPEDI (2: 2026: Florianópolis, Brasil).

CDU: 34

Conselho Nacional de Pesquisa
e Pós-Graduação em Direito Florianópolis
Santa Catarina – Brasil
www.conpedi.org.br

IX ENCONTRO VIRTUAL DO CONPEDI

DIREITO, GOVERNANÇA E NOVAS TECNOLOGIAS V

Apresentação

O Conselho Nacional de Pesquisa e Pós-graduação em Direito (CONPEDI), referência na difusão de pesquisas que abordam os novos fenômenos envolvendo o Direito, firma-se como espaço qualificado para apresentação de pesquisas produzidas na área das novas tecnologias em face do Direito.

O tema, recorrente nos artigos apresentados - o uso da inteligência artificial no âmbito jurídico, demonstra que a maior sociedade científica na área jurídica do Brasil se mantém atenta às inovações tecnológicas, ao avanço de diversas tecnologias, novas experiências no judiciário e nos escritórios de advocacia. Artigos, de alta relevância acadêmica e científica, destacam que o Direito não está imune a essas transformações e reiteram a necessidade de atenção aos desafios regulatórios do uso da inteligência artificial no mundo jurídico.

Diferentes pesquisas, apresentadas no Grupo de Trabalho 'Direito, Governança e Novas Tecnologias V', alertam para os efeitos deletérios da opacidade algorítmica,

referindo-se à falta de transparência sobre como sistemas de inteligência artificial e seus algoritmos produzem decisões, dentro e fora do judiciário. Nos textos, a governança

digital é debatida, com rigorosa visão científica, objetivando garantir o uso estratégico de tecnologias de informação e comunicação para gerir serviços, promover a segurança

de dados e integrar processos, de forma transparente, eficiente, com foco na participação ativa do cidadão na tomada de decisões.

algorítmico e tecnológico do sistema de justiça, englobando a investigação de cibercrimes, a digitalização de procedimentos policiais e judiciais, provas digitais e a

regulação de plataformas para combater conteúdos ilícitos.

Destarte, os artigos apresentados e os debates produzidos no Grupo de Trabalho ‘Direito, Governança e Novas Tecnologias V’ são essenciais para o

fortalecimento das diretrizes nacionais de regulação das novas tecnologias, pela formação de parâmetros normativos e procedimentais adequados à proteção dos

direitos humanos e fundamentais, à garantia da segurança da informação, à gestão de riscos e ao fortalecimento da transparência e do alinhamento legal no uso das novas

tecnologias no âmbito jurídico.

Por essas razões, as coordenadoras do GT convidam para a leitura do teor integral dos artigos, agradecendo a participação dos autores pesquisadores desta edição.

Vivianne Rigoldi/Jéssica Fachin/Rosane Porto

GOVERNANÇA DA INTELIGÊNCIA ARTIFICIAL: DA IMPRESCINDIBILIDADE DE EXPLICAR, TORNAR TRANSPARENTE E INTELIGÍVEL O FUNCIONAMENTO DO SISTEMA DA INTELIGÊNCIA ARTIFICIAL PARA A SUA REGULAÇÃO

ARTIFICIAL INTELLIGENCE GOVERNANCE: ON THE INDISPENSABILITY OF EXPLAINING, MAKING TRANSPARENT, AND ENSURING THE INTELLIGIBILITY OF ARTIFICIAL INTELLIGENCE SYSTEMS FOR THEIR REGULATION

Victória Alves Ruenreang [1]

Resumo

O presente artigo analisa a governança da inteligência artificial, com ênfase na imprescindibilidade dos princípios de explicabilidade, transparência e inteligibilidade para a construção de um ambiente regulatório capaz de assegurar previsibilidade e controle das decisões automatizadas. Parte-se do reconhecimento de que os sistemas de inteligência artificial, em razão de sua complexidade técnica, da dependência de grandes volumes de dados e da opacidade de seus processos decisórios, podem produzir vieses discriminatórios e resultados imprevistos, com potencial de violação de direitos fundamentais, especialmente os direitos da personalidade, a igualdade e a proteção de dados pessoais. A metodologia adotada baseia-se em pesquisa doutrinária e documental, com análise de casos concretos de falhas algorítmicas, bem como de instrumentos normativos nacionais e internacionais e diretrizes éticas voltadas à governança algorítmica. Examina-se, ainda, a interface entre esses princípios e o ordenamento jurídico brasileiro, especialmente a Constituição Federal, o Código de Defesa do Consumidor e a Lei Geral de Proteção de Dados. Conclui-se que a efetividade da regulação da inteligência artificial depende da integração entre princípios éticos e mecanismos jurídicos vinculantes, capazes de assegurar responsabilização, centralidade humana e proteção de direitos fundamentais.

Palavras-chave: Governança da inteligência artificial, Explicabilidade, Transparência, Inteligibilidade, Direitos fundamentais

violate fundamental rights, especially personality rights, equality, and personal data protection. The methodology is grounded in doctrinal and documentary research, including the analysis of concrete cases of algorithmic failures, as well as national and international regulatory instruments and ethical guidelines related to algorithmic governance. It also examines the interface between these principles and the Brazilian legal system, particularly the Federal Constitution, the Consumer Protection Code, and the General Data Protection Law. It concludes that the effectiveness of artificial intelligence regulation depends on the integration of ethical principles and binding legal mechanisms capable of ensuring accountability, human centrality, and the protection of fundamental rights.

Keywords/Palabras-claves/Mots-clés: Artificial intelligence governance, Explainability, Transparency, Intelligibility, Fundamental rights

1. Introdução

Apesar das bases de estudo da inteligência artificial terem surgido há milhares de anos,[1] junto com a filosofia,[2] a linguística, a psicologia e a biologia.[3] É apenas a partir da década de 40 e 50, sob a denominação de *machine intelligence*, que Alan Turing começa seus primeiros estudos na área, ao publicar seu artigo "*Computing Machinery and Intelligence*",[4] no qual estabelece fundamentos teóricos importantes para a discussão sobre a possibilidade de máquinas pensarem (*e.g.* Test de Turing).

Em 1956, passa a ser estudada oficialmente como disciplina científica e conhecida pelo seu nome atual (*Artificial Intelligence - AI*), após quatro pesquisadores norte-americanos[5] elaborarem um Projeto de Pesquisa de Verão na *Dartmouth College*, com base na conjectura de que todos os aspectos da aprendizagem ou qualquer outra característica da inteligência natural humana poderiam, em princípio, ser descritos com precisão suficiente para que uma máquina fosse capaz de simulá-los, permitindo a resolução de problemas tradicionalmente restritos aos seres humanos e o aprimoramento dessas capacidades.[6]

Desde então, são inúmeros os eventos que marcam a evolução da IA. Mas é no enredo clássico de homem contra máquina, onde a capacidade da IA de superar o homem fica nítida, principalmente em jogos que necessitam de elevado raciocínio e estratégia.[7] A cada vitória, os

---

[1] COPPIN, Ben. **Inteligência Artificial**. Rio de Janeiro, 2010. p. 3.

[2] "Sócrates alegava que poderia ser definido um algoritmo que descrevesse o comportamento de humanos e determinasse quando o comportamento de uma pessoa fosse bom ou mau. Isto nos leva a uma questão fundamental que tem sido formulada por filósofos e estudantes de Inteligência Artificial por muitos anos: há algo mais na mente do que simplesmente uma coleção de neurônios? Ou, dito de outra forma, se cada neurônio do cérebro humano fosse substituído por um dispositivo computacional equivalente, resultaria na mesma pessoa? Seria ele capaz de pensamento inteligente? " *In:* COPPIN, Ben. **Inteligência Artificial**. Rio de Janeiro, 2010. p. 10.

[3] Algumas pesquisas em IA se relacionam com os estudos desenvolvidos sobre o funcionamento dos neurônios do cérebro humano e com a psicologia cognitiva (produzir atos inteligentes com base no conhecimento).

[4] TURING, Alan M. Computing Machinery and Intelligence. **Oxford University Press on behalf of the Mind Association**, New Series, v. 59, n. 236, out. 1950), p. 433-460. Disponível em: https://phil415.pbworks.com/f/TuringComputing.pdf. Acesso em: 23 jun. 2023.

[5] John McCarthy (Dartmouth College), Marvin Minsky (Harvard University), Nathaniel Rochester (I.B.M. Corporation) e Claude Shannon (Bell Telephone Laboratories).

[6] "*We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.*" *In:* MCCARTHY, John; *et al*. **A proposal for the Dartmouth summer research project on artificial intelligence**. 1955. Disponível em: http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf. Acesso em: 20 jun. 2023.

[7] *e.g.* em 1996, o computador *Deep Blue*, da empresa IBM, vence o campeão mundial de xadrez da época. Em 2011, IBM Watson vence, em uma partida televisionada, os dois jogadores mais bem classificados do desafio de perguntas *Jeopardy!*. Em 2016, o sistema de inteligência artificial *AlphaGo*, da *Google DeepMind*, domina o antigo jogo de Go, ao derrotar um campeão mundial. Em 2017, o software *Libratus*, desenvolvida pelo *Carnegie*

sistemas de inteligência artificial são melhorados por seus desenvolvedores e passam a ter novas versões que solucionam problemas mais complexos em variadas áreas.

Na última década, a IA se tornou ponto relevante para qualquer tarefa intelectual, seja de maneira geral, especifica ou usual. Seu estudo floresceu em áreas de particular importância como o aprendizado de máquina; e seu avanço tem ocorrido em uma velocidade exponencial, que outas áreas do conhecimento, como o Direito, têm tido dificuldade em acompanhar. Nos tempos atuais, o objetivo do seu estudo "não é mais criar um robô tão inteligente quanto um humano, mas em vez disso usar algoritmos, heurísticas e metodologias baseadas nos modos pelos quais o cérebro humano soluciona problemas".[8]

Contudo, como qualquer outra atividade, a inteligência artificial também envolve riscos relevantes. Presume-se, e já há relatos nesse sentido, que os principais danos decorrentes dessa interação possam envolver lesões a direitos fundamentais previstos no artigo 5º da Constituição da República Federativa de 1988, como a inviolabilidade do direito à vida, à igualdade e à segurança, bem como, de forma especialmente sensível, aos direitos da personalidade, tais como privacidade, imagem, honra e proteção de dados pessoais. Estes riscos ganham relevo diante do fato de que os modelos e algoritmos mais avançados de sistemas de IA são treinados a partir de grandes volumes de dados, muitas vezes incluindo dados pessoais protegidos por legislação específica. Esse elemento funcional, somado à complexidade técnica e à imprevisibilidade desses sistemas, pode criar um ambiente propício a tratamentos indevidos ou abusivos de informações, caso não sejam estabelecidos limites legais, éticos e regulatórios.

Nesse contexto, a presente pesquisa tem como objetivo analisar os impactos da evolução da inteligência artificial e os riscos decorrentes de sua aplicação, com ênfase na possibilidade de violação de direitos fundamentais. A metodologia baseia-se na análise de casos reais envolvendo o desenvolvimento e uso de sistemas de IA, ao evidenciar falhas, vieses algorítmicos e resultados inesperados, além do exame de mecanismos de governança algorítmica e instrumentos normativos e éticos voltados à construção de sistemas mais confiáveis. Busca-se compreender como essas experiências e mecanismos contribuem para a formulação de parâmetros jurídicos adequados à complexidade tecnológica. Por fim, conclui-

Mellon University, vence quatro profissionais no pôquer. Em 2019, a OpenAI desenvolve a mão robótica, *Dactyl* que resolveu um cubo mágico sozinha; bem como o *OpenAI Five* que derrotou campeões mundiais no *e-sports game Dota 2*, um *jogo* multijogador de ação e estratégia em tempo real mais profundo já feito. Em 2020, o *Google DeepMind*, desenvolve o *Agent57*, o primeiro agente de aprendizado de reforço profundo a obter uma pontuação acima da linha de base humana em todos os 57 jogos *Atari 2600*. Em 2024, os modelos *AlphaProof* e *AlphaGeometry 2*, da *Google DeepMind*, resolvem problemas avançados de raciocínio na Olimpíada Internacional de Matemática, e alcançam um padrão de medalha de prata.

[8] COPPIN, Ben. **Inteligência Artificial**. Rio de Janeiro, 2010. p. 9.

se que os princípios éticos e as diretrizes de governança algorítmica são relevantes para mitigação de riscos, mas sua efetividade depende da integração com estruturas normativas capazes de assegurar a centralidade humana, a proteção de direitos fundamentais e a responsabilização pelos impactos da inteligência artificial.

2. Casos emblemáticos de falhas e vieses em sistemas de inteligência artificial e suas implicações para direitos fundamentais

A literatura traz alguns casos reais onde aplicações que, utilizam sistemas de IA, produziram resultados divergentes do esperado e acabaram por violar direitos fundamentais.

Em 2016, a Microsoft foi obrigada a retirar de circulação o *chatbot* Tay, desenvolvido para interações "casuais e brincalhonas" com jovens usuários na plataforma Twitter, após o sistema passar a publicar, em menos de 24 horas, conteúdos racistas, sexistas e xenofóbicos, mesmo tendo sido submetido a testes durante seu desenvolvimento. O caso evidencia a imprevisibilidade desses sistemas e a relevância da qualidade dos dados de entrada (*input*) para a geração de resultados (*outputs*). Tay não foi o primeiro *chatbot* da empresa, já que o XiaoIce, utilizado na China por milhões de usuários, havia obtido êxito, o que motivou a adaptação do modelo para outro contexto cultural; contudo, essa transposição resultou em efeitos opostos aos inicialmente observados.[9]

Em 2018, uma pesquisa desenvolvida pelo MIT e pela Universidade de Stanford, constatou que três programas de inteligência artificial, lançados comercialmente por grandes empresas e que já estavam em circulação, demonstraram ter preconceito de gênero e de tipo de pele no seu processo de inferir. Nos experimentos, "as taxas de erro [não reconhecimento facial] dos três programas na determinação do gênero de homens de pele clara nunca foram piores do que 0,8%. Para mulheres de pele mais escura, no entanto, as taxas de erro dispararam para mais de 20% em um caso e mais de 34% nos outros dois [tradução nossa]".[10]

---

[9] "*As many of you know by now, on Wednesday we launched a chatbot called Tay. We are deeply sorry for the unintended offensive and hurtful tweets from Tay, which do not represent who we are or what we stand for, nor how we designed Tay. Tay is now offline and we'll look to bring Tay back only when we are confident we can better anticipate malicious intent that conflicts with our principles and values*". *In:* LEE, Peter. Learning from Tay's introduction. **Official Microsoft Blog**. 2016. Disponível em: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/. Acesso em: 22 out. 2024.

[10] Texto original: "*In the researchers' experiments, the three programs' error rates in determining the gender of light-skinned men were never worse than 0.8 percent. For darker-skinned women, however, the error rates ballooned — to more than 20 percent in one case and more than 34 percent in the other two*". *In:* HARDESTY, Larry. Study finds gender and skin-type bias in commercial artificial-intelligence systems. **MIT News**. 2018. Disponível em: https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212. Acesso em: 22 out. 2024.

Os estudos do MIT demonstram que:

> As descobertas levantam questões sobre como as redes neurais de hoje, que aprendem a executar tarefas computacionais procurando padrões em enormes conjuntos de dados, são treinadas e avaliadas. Por exemplo, de acordo com o artigo, pesquisadores de uma grande empresa de tecnologia dos EUA alegaram uma taxa de precisão de mais de 97% para um sistema de reconhecimento facial que eles projetaram. Mas o conjunto de dados usado para avaliar seu desempenho era mais de 77% masculino e mais de 83% branco [tradução nossa].[11]

O problema de sistemas de inteligência artificial enviesados constitui um dos principais desafios contemporâneos do Direito, sobretudo devido à coleta massiva de dados em diferentes contextos, que sustenta seu funcionamento.

No Brasil a proteção dos dados pessoas, nova espécie da categoria jurídica dos direitos da personalidade,[12] está previsto como um direito fundamental na Constituição da República Federativa do Brasil de 1988 (CRFB 88), no seu artigo 5º, inciso LXXIX, onde "é assegurado, nos termos da lei, o direito à proteção dos dados pessoais, inclusive nos meios digitais".[13] A referida lei mencionada na Constituição, é a Lei nº 13.709, de 14 de agosto de 2018, conhecida também como Lei Geral de Proteção de Dados (LGPD), que dispõe sobre o "tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural".[14]

Entretanto, "todos esses dados pessoais que, antes das leis específicas de proteção, careciam de um marco regulatório [próprio], viram-se diante de uma outra problemática: a [ausência de] regulação das ferramentas tecnológicas empregadas nesse meio [*e.g.* o sistema de IA] ".[15] Mas mesmo diante de uma possível lacuna normativa, em muitos pontos a LGPD traz

---

[11] Texto original: "*The findings raise questions about how today's neural networks, which learn to perform computational tasks by looking for patterns in huge data sets, are trained and evaluated. For instance, according to the paper, researchers at a major U.S. technology company claimed an accuracy rate of more than 97 percent for a face-recognition system they'd designed. But the data set used to assess its performance was more than 77 percent male and more than 83 percent white*". *In:* HARDESTY, Larry. Study finds gender and skin-type bias in commercial artificial-intelligence systems. **MIT News**. 2018. Disponível em: https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212. Acesso em: 22 out. 2024.

[12] BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, p. 94-132. *E-book.*

[13] BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 mai. 2024.

[14] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 3 mai. 2024.

[15] FERNANDES, Rafael Gonçalves; OLIVEIRA, Liziane Paixão Silva. A regulação do agir decisório disruptivo no judiciário brasileiro e a observância do princípio da precaução: juiz natural ou "juiz artificial"? **Revista Opinião Jurídica**, Fortaleza, v. 19, n. 30, p. 99-100, jan./abr. 2021. Disponível em: https://periodicos.unichristus.edu.br/opiniaojuridica/article/view/3446. Acesso em: 05 mai. 2023.

um norte dos possíveis problemas e suas possíveis soluções no campo de IA, quando esta utiliza dados pessoais. A partir de uma análise jurídica dos casos, destacam-se duas principais categorias de riscos: a violação dos direitos da personalidade e a violação do direito à igualdade.

No contexto tecnológico, a violação dos direitos da personalidade previstos no artigo 5º, inciso X, da Constituição Federal, especialmente da intimidade e da privacidade, transcende o tradicional direito de ser deixado em paz (*the right to be let alone*), formulado por Samuel D. Warren e Louis D. Brandeis. [16] As barreiras físicas da vida privada já não isolam o indivíduo do mundo, tornando o compartilhamento de dados pessoais frequentemente necessário para a vida em sociedade. Nessa realidade, a proteção da intimidade e da privacidade passou a envolver não apenas o direito de isolamento, mas também o controle, a veracidade e o tratamento adequado dos dados pessoais compartilhados. A tutela jurídica busca assegurar que essas informações sejam tratadas conforme a lei, protegendo direitos fundamentais como intimidade, privacidade, honra e imagem, e garantindo o livre desenvolvimento da personalidade.[17] Nesse contexto, destaca-se a autodeterminação informativa, prevista no artigo 2º, inciso II, da LGPD, que assegura ao titular o direito de controlar seus dados pessoais.[18]

O segundo risco refere-se à violação do direito à igualdade, previsto no art. 5º, caput, da Constituição Federal. Para que as pessoas possam desenvolver livremente sua personalidade, não podem ser discriminadas. Embora a discriminação algorítmica seja frequentemente associada apenas ao ambiente digital, ela reflete vieses historicamente presentes na sociedade, agora amplificados pelo volume e pela velocidade de processamento dos sistemas de IA. Como o compartilhamento de dados digitais se tornou necessário para a vida em sociedade, é essencial evitar que esses vieses sejam replicados no tratamento das informações. Alguns dados, considerados sensíveis por lei, como origem racial ou étnica, convicção religiosa ou opinião política, apresentam maior potencial de gerar discriminação, mas mesmo dados inicialmente não sensíveis podem se tornar problemáticos dependendo do contexto em que são utilizados.

É em consonância com as três dimensões de igualdade (formal, material e como reconhecimento) [19] contempladas na Constituição de 1988, que estão previstas nos direitos

---

[16] WARREN, Samuel D.; BRANDEIS, Louis D. The right to privacy**. Harvard Law Review**, Vol. 4, n. 5, 1890, p. 193-220. Disponível em: https://www.jstor.org/stable/1321160. Acesso em: 07 de ago de 2023.

[17] *BESSA, Leonardo Roscoe.* A Lei Geral de Proteção de Dados e o direito à autodeterminação informativa. **Conjur**. 2020. Disponível em: https://www.conjur.com.br/2020-out-26/leonardo-bessa-lgpd-direito-autodeterminacao-informativa. Acesso em: 07 de ago de 2023.

[18] BESSA, Leonardo Roscoe. A Lei Geral de Proteção de Dados e o direito à autodeterminação informativa. **Conjur**. 2020. Disponível em: https://www.conjur.com.br/2020-out-26/leonardo-bessa-lgpd-direito-autodeterminacao-informativa. Acesso em: 07 de ago de 2023.

[19] "A igualdade constitui um direito fundamental e integra o conteúdo essencial da ideia de democracia. Da dignidade humana resulta que todas as pessoas são fins em si mesmas, possuem o mesmo valor e merecem, por essa razão, igual respeito e consideração. A igualdade veda a hierarquização dos indivíduos e as desequiparações

fundamentais (artigo 5º, *caput*, CRFB 88 – igualdade formal)[20] e nos objetivos fundamentais da República Federativa do Brasil (incisos I e III do artigo 3º da CRFB 88 - igualdade material; inciso IV do artigo 3º da CRFB 88 - igualdade como reconhecimento).[21] Que o artigo 6º, inciso IX, da LGPD, prevê o princípio da não descriminação: "impossibilidade de realização de tratamento para fins discriminatórios ilícitos ou abusivos".[22] Ao colaborar para a prevenção de práticas discriminatórias, a LGPD não só segue os preceitos constitucionais como também inova, ao impedir que o dado seja processado para aquela finalidade antes mesmo que o seu tratamento indevido ocorra.[23]

Conforme a interpretação do referido artigo, práticas discriminatórias sem finalidade ilícita ou abusiva são permitidas. Onde a ilicitude decorre da violação da igualdade formal, enquanto a abusividade, embora não definida expressamente, refere-se a atos lícitos que se tornam inadequados em determinados contextos, podendo levantar questões éticas.[24]

Apesar do uso de decisões automatizadas se justificar pela tentativa de superar falhas e vieses humanos, a ausência de subjetividade não garante neutralidade dos sistemas de IA.[25]

---

infundadas, mas impõe a neutralização das injustiças históricas, econômicas e sociais, bem como o respeito à diferença. Em torno de sua maior ou menor centralidade nos arranjos institucionais, bem como no papel do Estado na sua promoção, dividiram-se as principais ideologias e correntes políticas dos últimos séculos. No mundo contemporâneo, a igualdade se expressa particularmente em três dimensões: a igualdade formal, que funciona como proteção contra a existência de privilégios e tratamentos discriminatórios; a igualdade material, que corresponde às demandas por redistribuição de poder riqueza e bem estar social; e a igualdade como reconhecimento, significando o respeito devido às minorias, sua identidade e suas diferenças, sejam raciais, religiosas, sexuais ou quaisquer outras". *In:* BARROSO, Luís Roberto; OSÓRIO, Aline. "Sabe com quem está falando? ": Notas sobre o princípio da igualdade no Brasil contemporâneo. **Revista Direito e Práxis**, v. 7, n. 1, p. 207-208, 2016. Disponível em: https://www.e-publicacoes.uerj.br/index.php/revistaceaju/article/view/21094/15886. Acesso em: 01 abr. 2023.

[20] "Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade (...)". *In*: BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 mai. 2024.

[21] "Art. 3º Constituem objetivos fundamentais da República Federativa do Brasil: I - construir uma sociedade livre, justa e solidária; (...); III - erradicar a pobreza e a marginalização e reduzir as desigualdades sociais e regionais; IV - promover o bem de todos, sem preconceitos de origem, raça, sexo, cor, idade e quaisquer outras formas de discriminação". *In*: BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 mai. 2024.

[22] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados (LGPD). Brasília, DF: Diário Oficial da União, 2018.

[23] MENDES, Laura Schertel; MATTIUZO, Marcela; FUJIMOTO, Mônica Tiemy. Discriminação algorítmica à luza da Lei Geral de Proteção de Dados. p. 421-518. *In:* MENDES, Laura; DONEDA, Danilo; SARLET, Ingo Wolfgang *et al*. (Coord.) **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

[24] MATTIUZZO, Marcela. Discriminação algorítmica: reflexões no contexto da Lei Geral de Proteção de Dados Pessoais, p. 117-126. *In:* MENDES, Laura; DONEDA, Danilo; CUEVA, Ricardo Villas Bôas. **Lei Geral de Proteção de Dados (Lei 13.709/2018): a caminho da efetividade**. São Paulo: Revista dos Tribunais, 2020.

[25] REQUIÃO, Maurício; COSTA, Diego. Discriminação algorítmica: ações afirmativas como estratégia de combate. **civilistica.com**, [S. l.], v. 11, n. 3, p. 3-4, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 15 mar. 2024.

A discriminação algorítmica surge de duas formas principais: primeiro, quando os algoritmos refletem os preconceitos humanos, conscientes ou não, embutidos desde a programação.[26] O exemplo dos estudos realizados pelo MIT com os programas de reconhecimento facial, retratam bem isso. Onde o seu desenvolvedor acaba por contaminar o programa com os seus preconceitos, seja de maneira intencional ou não, ao utilizar dados incompletos, pouco representativos ou tendenciosos no seu aprendizado.

A segunda forma de viés ocorre quando os algoritmos entram em contato com bases de dados já contaminadas por preconceitos, aprendendo a reproduzir discriminações existentes. Nessa situação, mesmo sem influência direta do programador, vieses algorítmicos podem se infiltrar quando os dados fornecidos aos sistemas de IA são de baixa qualidade ou confiabilidade, refletindo desigualdades presentes na própria sociedade. O caso do Chatbot Tay ilustra bem esse fenômeno: ao aprender com dados do Twitter, frequentemente usados para disseminar discurso de ódio, o algoritmo reproduziu comportamentos discriminatórios.[27]

Como o sistema de IA não tem discernimento próprio (subjetividade) para ter seus próprios valores e convicções a partir dessas conexões e informações que lhes são dadas. Ele passa então a reproduzir, de maneira generalizada, vieses já preexistentes em seus programadores e desenvolvedores ou, presentes nos dados utilizados como base para o seu aprendizado. Questão que se agrava pela ausência de transparência, explicabilidade e inteligibilidade do processo de inferir dos sistemas de IA, mas que pode ser evitada pela promoção desses princípios (governança algorítmica).

3. Governança algorítmica e proteção de direitos fundamentais na inteligência artificial

Para compatibilizar, na esfera legal, a natureza imprevisível dos sistemas de IA com a previsibilidade e a confiança exigidas pelas normas, pesquisadores têm buscado meios técnicos de compreender seus processos decisórios. Nesse contexto, surgiu o conceito de governança algorítmica, baseada em princípios que orientam o desenvolvimento de algoritmos e modelos de IA de forma responsável e confiável. Embora recente, diversas organizações já propuseram princípios próprios para uma IA confiável, que se assemelham e se diferenciam, tornando o

[26] REQUIÃO, Maurício; COSTA, Diego. Discriminação algorítmica: ações afirmativas como estratégia de combate. **civilistica.com**, [S. l.], v. 11, n. 3, p. 4, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 15 mar. 2024.

[27] REQUIÃO, Maurício; COSTA, Diego. Discriminação algorítmica: ações afirmativas como estratégia de combate. **civilistica.com**, [S. l.], v. 11, n. 3, p. 4, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 15 mar. 2024.

campo ainda vasto e em consolidação. Diante disso, seguem alguns documentos que representam medidas propostas para o desenvolvimento consciente da inteligência artificial:

1. *Fairness, Accountability and Transparency in Machine Learning Organization - FAT/ML*;[28]
2. *"The Code" - Association for Computing Machinery Code of Ethics and Professional Conduct*;[29]
3. *Asimolar AI principles - Future of Life Institute*;[30]
4. *OECD AI Principles* (OECD/LEGAL/0449);[31] [32]
5. *The Montreal Declaration for the Responsible development of Artificial Intelligence*;[33]
6. *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*;[34]
7. *Statement on Artificial Intelligence, robotics and 'autonomous' systems: European Commission's European group on ethics in Science and new Technologies*;[35]
8. *AI in the UK: ready willing and able*;[36]
9. Orientações éticas para uma IA de confiança da GPAN IA (Grupo Independente de Peritos de Alto Nível sobre a Inteligência Artificial, criado pela Comissão Europeia em julho de 2018);[37]
10. Estratégia Brasileira de Inteligência Artificial - EBIA, Instituída pela Portaria MCTI nº 4.617, de 6 de abril de 2021, alterada pela Portaria MCTI nº 4.979, de 13 de julho de 2021.[38]

---

[28] DIAKOPOULOS, Nicholas *et al*. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. **FAT/ML**. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 20 jul. 2024.

[29] ACM Code of Ethics and Professional Conduct. **Association for Computing Machinery**. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 20 jul. 2024.

[30] THE ASIMOLAR AI Principles. **The future of life**. 2017. Disponível em: https://futureoflife.org/open-letter/ai-principles/. Acesso em: 20 jul. 2024.

[31] OECD AI Principles overview. **OECD.AI Policy Observatory**. 2024. Disponível em: https://oecd.ai/en/ai-principles. Acesso em: 20 jul. 2024.

[32] OECD/LEGAL/0449. Recommendation of the Council on Artificial Intelligence. **OECD Legal Instruments**. 2024. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 20 jul. 2024.

[33] Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial 2018. **Declaração de Montreal pela IA responsável**. 2018. Disponível em: https://declarationmontreal-iaresponsable.com/. Acesso em: 20 jul. 2024.

[34] Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. **IEEE Advancing Technology for Humanity**. 2019. Disponível em: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf. Acesso em: 20 jul. 2024.

[35] EUROPEAN Group on Ethics in Science and New Technologies. Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. **European Commission**. mar. 2018. Disponível em: https://www.unapcict.org/sites/default/files/2019-01/EC_AI-%20Robotics-%20and%20Autonomous%20Systems.pdf. Acesso em: 20 jul. 2024.

[36] HOUSE of Lords. AI in the UK: ready, willing and able? **Select Committee on Artificial Intelligence** abr. 2019. Disponível em: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. Acesso em: 20 jul. 2024.

[37]High-Level Expert Group on Artificial Intelligence (AI HLEG). Ethics guidelines for trustworthy AI. **European Commission**. Abr. 2019. Disponível em: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf. Acesso em: 20 jul. 2024.

[38] BRASIL, Ministério da Ciência, Tecnologia e Inovações. **Inteligência Artificial**. Brasília, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial. Acesso em: 20 de ago. de 2023.

Após análise das fontes mencionadas, quatro princípios se destacam para conferir previsibilidade aos sistemas de IA e garantir a proteção dos direitos fundamentais: explicabilidade, transparência, inteligibilidade e *accountability* (prestação de contas). Por se complementarem, nem sempre aparecem sob a mesma nomenclatura, podendo também ser combinados em um único princípio. Por isso, serão apresentados, de forma geral, os princípios e sua fundamentação, oferecendo um panorama geral do tema.

Para Finale Doshi-Velez *et al*, a ideia por trás do princípio da explicabilidade, desenvolvido especificamente na seara da IA, quando aplicado a um processo decisório "refere-se a um conjunto de razões abstraídas ou justificativas para um resultado específico, não uma descrição do processo de tomada de decisão em geral [tradução nossa]".[39] Onde "o termo 'explicação' significa informações interpretáveis por humanos sobre a lógica pela qual um tomador de decisão tomou um conjunto específico de informações e chegou a uma conclusão específica [tradução nossa]".[40] Ao "garantir que as decisões algorítmicas, bem como quaisquer dados que as orientem, possam ser explicados aos usuários finais e outras partes interessadas em termos não técnicos [tradução nossa]".[41]

"É importante notar que explicabilidade não é o mesmo que transparência, na medida em que ser capaz de entender o processo por meio do qual uma decisão foi tomada não é o mesmo que conhecer todos os passos tomados para se atingir aquela decisão".[42] O princípio da transparência, parte do pressuposto de que o sistema de IA não pode estra coberto pelo segredo.[43] Onde a "honestidade é um componente essencial da confiabilidade [tradução

---

[39] Texto original: "*In this paper, when we talk about an explanation for a decision, we mean a set of abstracted reasons or justifications for a particular outcome, not a description of the decision-making process in general*". *In:* DOSHI-VELEZ, Finale *et al*. Accountability of AI Under the Law: The Role of Explanation. **Harvard Public Law**, p. 4, 2017. Disponível em: https://www.researchgate.net/publication/320867322_Accountability_of_AI_Under_the_Law_The_Role_of_Explanation. Acesso em 06 jul. 2024.

[40] Texto original: "*More specifically, we define the term 'explanation' to mean human-interpretable information about the logic by which a decision-maker took a particular set of inputs and reached a particular conclusion". In:* DOSHI-VELEZ, Finale *et al*. Accountability of AI Under the Law: The Role of Explanation. **Harvard Public Law**, p. 4, 2017. Disponível em: https://www.researchgate.net/publication/320867322_Accountability_of_AI_Under_the_Law_The_Role_of_Explanation. Acesso em 06 jul. 2024.

[41] Texto original: "*Ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms*". DIAKOPOULOS, Nicholas *et al*. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. **FAT/ML**. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 20 jul. 2024.

[42] MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **RDU**, Porto Alegre, Volume 16, n. 90, p. 18, nov-dez 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766/Schertel%20Mendes%3B%20Mattiuzzo%2C%202019. Acesso em: 06 jul. 2024.

[43] FACHIN, Zulmar; FACHIN, Jéssica; SILVA, Deise Marcelino da. Princípios de inteligência artificial. **Constituição, Economia e Desenvolvimento: Revistada Academia Brasileira de Direito Constitucional**.

nossa]",[44] pois "o poder invisível se forma e se organiza não somente para combater o poder público, mas também para tirar benefícios ilícitos e extrair dele vantagens que não seriam permitidas por uma ação à luz do dia".[45] A clareza é "condição necessária para [que] tais agentes inteligentes possam argumentar e explicar as decisões por eles tomadas".[46]

O princípio da inteligibilidade é incorporado ao princípio da explicabilidade em algumas fontes, onde "deve-se destacar que, ao se desenvolverem aplicações de IA, há um destinatário final que fará uso do produto desenvolvido, logo, este consumidor deve possuir consciência acerca do funcionamento da IA".[47] Parte do pressuposto de que a compreensão do seu funcionamento não deve ficar retida a um grupo de especialistas no assunto e por desenvolvedores de aplicações de IA, mas sim a todos. Ele não deve ser apenas passível de ser explicado com também deve ser entendido (inteligível) pela sociedade, foco principal da interação entre homem e máquina. Onde os profissionais de computação devem compartilhar de maneira clara, respeitosa e acolhedora, o seu conhecimento técnico com o público, ao promover a conscientização sobre a computação e suas consequências, bem como encorajar a sua compreensão.[48]

A recomendação da OCDE traz de forma expressa o princípio da transparência e da explicabilidade, assim como a ideia do princípio da inteligibilidade, conforme a seguir:

> **1.3. Transparência e explicabilidade (e inteligibilidade)**
> Os intervenientes na IA devem comprometer-se com a transparência e a divulgação responsável em relação aos sistemas de IA. Para o efeito, devem fornecer informações significativas, adequadas ao contexto e consistentes com o estado da arte:
> i. promover uma compreensão geral dos sistemas de IA, incluindo as suas capacidades e limitações,
> ii. conscientizar as partes interessadas sobre suas interações com os sistemas de IA, inclusive no local de trabalho,
> iii. sempre que viável e útil, fornecer informações simples e fáceis de entender sobre as fontes de dados/entradas, fatores, processos e/ou lógica que levaram à previsão, conteúdo, recomendação ou decisão, para permitir que as pessoas afetadas por um Sistema de IA para entender o resultado e,

Curitiba, 2022, vol. 14, n. 26, p. 371, jan./jul.,2022. Disponível em: https://abdconstojs.com.br/index.php/revista/article/view/434/292. Acesso em: 06 jul 2024.

[44] Texto original: "Honesty is an essential component of trustworthiness. A computing professional should be transparent and provide full disclosure of all pertinent system capabilities, limitations, and potential problems to the appropriate parties". *In:* ACM Code of Ethics and Professional Conduct. **Association for Computing Machinery**. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 20 jul. 2024.

[45] BOBBIO, Norberto. **Democracia e Segredo**. São Paulo: UNESP, 2015. p. 33.

[46] SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. **Revista USP**. 2021. Disponível em: <https://www.revistas.usp.br/eav/article/view/185024/171209>. Acesso em: 06 jul. 2024.

[47] LIMA, Cíntia Rosa Pereira de; OLIVEIRA, Cristina Godoy Bernardo de; RUIZ, Evandro Eduardo Seron. Inteligência artificial e personalidade jurídica: aspectos controvertidos. p. 124. *In:* BARBOSA, Mafalda Miranda; et al (org.). **Direito digital e inteligência artificial:** diálogos entre Brasil e Europa. Indaiatuba, SP: Editora Foco, 2021.

[48] ACM Code of Ethics and Professional Conduct. **Association for Computing Machinery**. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 20 jul. 2024.

> iv.fornecer informações que permitam às pessoas afetadas negativamente por um sistema de IA contestar os seus resultados [tradução nossa].[49]

Extrai-se do ordenamento vigente, a ideia central dos princípios da transparência, explicabilidade e inteligibilidade, mesmo sem uma lei especifica que os aplique diretamente ao uso de sistemas de IA. Como uma prerrogativa essencial reconhecida ao titular dos dados e ao consumidor desta tecnologia.[50] Prevista de maneira geral nas relações de consumo, que também envolvem grande parte das relações entre homem-máquina. O artigo 6º do Código de Defesa do Consumidor traz que "são direitos básicos do consumidor: (...) III - a informação adequada e clara sobre os diferentes produtos e serviços, com especificação correta de quantidade, características, composição, qualidade, tributos incidentes e preço, bem como sobre os riscos que apresentem".[51] Onde sua insuficiência e inadequação podem resultar na responsabilização objetiva de seus responsáveis legais (artigos 12 a 17 do CDC).

A Lei Geral de Proteção de Dados ao "oferecer mecanismos para minimizar os riscos desencadeados pelo crescente uso de algoritmos para realização de julgamentos e avaliações das pessoas, referentes aos quais não são revelados as razões e processos de decisão".[52] Prevê em seu artigo 20, o direito de revisão. Seu objetivo inicial (revisar uma decisão automatizada) ainda subsiste no *caput* do seu artigo, assim como o princípio da explicabilidade no seu § 1º, que dá ao titular dos dados o direito de saber como seus dados foram utilizados:

> Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

---

[49] Texto original: "**1.3. Transparency and explainability:** AI Actors should commit to transparency and responsible disclosure regarding AI systems. To this end, they should provide meaningful information, appropriate to the context, and consistent with the state of art: i. to foster a general understanding of AI systems, including their capabilities and limitations; ii. to make stakeholders aware of their interactions with AI systems, including in the workplace; iii. where feasible and useful, to provide plain and easy-to-understand information on the sources of data/input, factors, processes and/or logic that led to the prediction, content, recommendation or decision, to enable those affected by an AI system to understand the output, and; iv. to provide information that enable those adversely affected by an AI system to challenge its output". *In:* OECD/LEGAL/0449. Recommendation of the Council on Artificial Intelligence. **OECD Legal Instruments**. 2024. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 20 jul. 2024.

[50] GODOY, Claudio Luiz Bueno de. A intervenção humana no procedimento de explicação e de revisão de decisões automatizadas. p. 287-309. *In:* CHINELLATO, Silmara J. de A.; TOMASEVICIUS FILHO, Eduardo. **Inteligência Artificial**: Visões interdisciplinares e internacionais. São Paulo: Almedina, 2023.

[51] BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, [1990]. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 20 out. 2024.

[52] SILVA, Priscila; MEDEIROS, Juliana. A polêmica da revisão (humana) sobre decisões automatizada. ITS Rio, Rio de Janeiro, 2019. Disponível em: https://feed.itsrio.org/a-pol%C3%AAmica-da-revis%C3%A3o-humana-sobre-decis%C3%B5es-automatizadas-a81592886345. Acesso em: 17 abr. 2023.

> § 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial.
>
> § 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais.[53]

Contudo, mais completa a norma estaria com a presença do seu § 3º, que previa a revisão por pessoa natural no caso previsto no *caput* do artigo 20. Uma garantia extra que "atenderia ao princípio da transparência e colocaria a Legislação Brasileira no mesmo patamar de outras legislações internacionais que garantem a eficácia deste direito através da intervenção humana".[54] Entretanto, por razões de interesse público, o referido parágrafo foi vetado pelo Presidente da República. [55] Esta decisão acabou por prejudicar "o seu objetivo de proteger o sujeito contra ocasiões que o coloque em uma posição de vulnerabilidade em função de decisões potencialmente arbitrárias sem quaisquer justificativas ou possibilidade de recurso". [56] Uma vez que, hoje em dia, esta revisão será feita por um outro sistema autônomo.

Os três princípios mencionados (explicabilidade, transparência e inteligibilidade), servem de base para se desenvolver uma IA confiável. Onde "é importante existir a compreensão do seu funcionamento, de seu desenvolvimento e das tarefas a serem realizadas, pois a confiança [elemento básico das relações sociais] apenas pode ser estabelecida quando há nitidez quanto ao funcionamento do que se utiliza".[57] Esses princípios corroboram para facilitar

---

[53] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, [2018]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 20 out. 2024.

[54] SIQUEIRA, Dirceu Pereira; MORAIS, Fausto Santos de; TENA, Lucimara Plaza. O papel emacipador do direito em um contexto de linhas abissais e algoritmos. **Pensar: Revista de Ciências Jurídicas**, v. 7, n. 1, p. 9, 2022. Disponível em: https://ojs.unifor.br/rpen/article/view/12058/6780. Acesso em: 27 set. 2024.

[55] "Ouvidos, os Ministérios da Economia, da Ciência, Tecnologia, Inovações e Comunicações, a Controladoria-Geral da União e o Banco Central do Brasil manifestaram-se pelo veto ao seguinte dispositivo: (...) '§ 3º A revisão de que trata o caput deste artigo deverá ser realizada por pessoa natural, conforme previsto em regulamentação da autoridade nacional, que levará em consideração a natureza e o porte da entidade ou o volume de operações de tratamento de dados'. Razões do veto: 'A propositura legislativa, ao dispor que toda e qualquer decisão baseada unicamente no tratamento automatizado seja suscetível de revisão humana, contraria o interesse público, tendo em vista que tal exigência inviabilizará os modelos atuais de planos de negócios de muitas empresas, notadamente das startups, bem como impacta na análise de risco de crédito e de novos modelos de negócios de instituições financeiras, gerando efeito negativo na oferta de crédito aos consumidores, tanto no que diz respeito à qualidade das garantias, ao volume de crédito contratado e à composição de preços, com reflexos, ainda, nos índices de inflação e na condução da política monetária' ". *In:* BRASIL. **Mensagem nº 288, de 8 de julho de 2019. Brasília: Presidência da República, 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Msg/VEP/VEP-288.htm. Acesso em: 17 abr. 2023.**

[56] SILVA, Priscila; MEDEIROS, Juliana. A polêmica da revisão (humana) sobre decisões automatizada. **ITS Rio**, Rio de Janeiro, 2019. Disponível em: https://feed.itsrio.org/a-pol%C3%AAmica-da-revis%C3%A3o-humana-sobre-decis%C3%B5es-automatizadas-a81592886345. Acesso em: 17 abr. 2023.

[57] LIMA, Cíntia Rosa Pereira de; OLIVEIRA, Cristina Godoy Bernardo de; RUIZ, Evandro Eduardo Seron. Inteligência artificial e personalidade jurídica: aspectos controvertidos. p. 124. *In:* BARBOSA, Mafalda Miranda;

a aplicação do princípio da accountability (prestação de contas). Ao "permitir que terceiros interessados investiguem, entendam e revisem o comportamento do algoritmo por meio da divulgação de informações que permitem monitoramento, verificação ou crítica [tradução nossa]".[58] Pois, não basta apenas implementar medidas adequadas para cumprir a lei, é necessário também comprovar a sua conformidade com esta, quando for solicitado.

Como também auxiliam na responsabilização, ao "disponibilizar vias de reparação externamente visíveis para efeitos individuais ou sociais adversos de um sistema de decisão algorítmica e designar uma função interna para a pessoa responsável pela solução oportuna de tais problemas [tradução nossa]".[59] A fim de mitigar quaisquer potenciais danos e impactos sociais negativos, ao partir da premissa de que "o desenvolvimento e o uso de (...) [sistemas de IA] não devem contribuir para a desresponsabilização dos seres humanos quando uma decisão vier a ser tomada",[60] pois, "designers e construtores de sistemas avançados de IA são partes interessadas nas implicações morais da sua utilização, indevida ou não, com a responsabilidade e a oportunidade de molda-las [tradução nossa]".[61] Para Nicholas Diakopoulos *et al*:

> Algoritmos e os dados que os impulsionam são projetados e criados por pessoas - Há sempre um humano responsável final pelas decisões tomadas ou informadas por um algoritmo. "O algoritmo fez isso" não é uma desculpa aceitável se sistemas algorítmicos cometem erros ou têm consequências indesejadas, incluindo processos de aprendizado de máquina [tradução nossa].[62]

et al (org.). **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Editora Foco, 2021.

[58] Texto original: "*Enable interested third parties to probe, understand, and review the behavior of the algorithm through disclosure of information that enables monitoring, checking, or criticism, including through provision of detailed documentation, technically suitable APIs, and permissive terms of use*". *In:* DIAKOPOULOS, Nicholas *et al*. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. **FAT/ML**. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 20 jul. 2024.

[59] Texto original: "*Make available externally visible avenues of redress for adverse individual or societal effects of an algorithmic decision system, and designate an internal role for the person who is responsible for the timely remedy of such issues*". *In:* DIAKOPOULOS, Nicholas *et al*. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. **FAT/ML**. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 20 jul. 2024.

[60] Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial 2018. **Declaração de Montreal pela IA responsável**. p. 16. 2018. Disponível em: https://declarationmontreal-iaresponsable.com/. Acesso em: 20 jul. 2024.

[61] Texto original: "*Responsibility: Designers and builders of advanced AI systems are stakeholders in the moral implications of their use, misuse, and actions, with a responsibility and opportunity to shape those implications". In:* THE ASIMOLAR AI Principles. **The future of life**. 2017. Disponível em: https://futureoflife.org/open-letter/ai-principles/. Acesso em: 20 jul. 2024.

[62] Texto original: "*Algorithms and the data that drive them are designed and created by people -- There is always a human ultimately responsible for decisions made or informed by an algorithm. "The algorithm did it" is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes*". *In:* DIAKOPOULOS, Nicholas et al. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. **FAT/ML**. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 20 jul. 2024.

"A IA Responsável (ou IA Ética, ou IA Confiável) não é, como alguns podem afirmar, uma forma de dar as máquinas algum tipo de 'responsabilidade' por suas ações e decisões e, no processo, descarregar pessoas e organizações de sua responsabilidade [tradução nossa]".[63] "(...) [O sistema de IA] não pode encarar-se como o garante do humano por detrás de si, na medida em que não é responsável pela pessoa, mas inversamente, o homem é responsável pelo uso do ente dotado de inteligência artificial".[64] "(...) Desenvolvimento responsável e o uso da IA exige mais responsabilidade e mais responsabilização por parte das pessoas e organizações envolvidos: para as decisões e ações das aplicações de IA, e para a sua própria decisão de usar IA em um determinado contexto de aplicativo [tradução nossa]".[65]

Na resolução da OCDE, a responsabilidade proposta segue uma postura de prevenção (antecipa o dano), onde seus intervenientes devem agir de modo responsável em todas as fases.

> **1.5. Responsabilidade**
> a) Os intervenientes na IA devem ser responsáveis pelo bom funcionamento dos sistemas de IA e pelo respeito dos princípios acima referidos, com base nas suas funções, no contexto e em consonância com o estado da arte.
> b) Para o efeito, os intervenientes na IA devem garantir a rastreabilidade, nomeadamente em relação aos conjuntos de dados, processos e decisões tomadas durante o ciclo de vida do sistema de IA, para permitir a análise dos resultados do sistema de IA e das respostas à investigação, adequadas ao contexto e consistentes com o estado da arte.
> c) Os intervenientes na IA devem, com base nas suas funções, no contexto e na sua capacidade de agir, aplicar uma abordagem sistemática de gestão de riscos a cada fase do ciclo de vida do sistema de IA de forma contínua e adotar uma conduta empresarial responsável para abordar os riscos relacionados com a IA sistemas de IA, nomeadamente, se for caso disso, através da cooperação entre diferentes intervenientes na IA, fornecedores de conhecimentos e recursos de IA, utilizadores de sistemas de IA e outras partes interessadas. Os riscos incluem aqueles relacionados com preconceitos prejudiciais, direitos humanos, incluindo segurança, proteção e privacidade, bem como direitos laborais e de propriedade intelectual [tradução nossa].[66]

---

[63] Texto original: "*Responsible AI (or Ethical AI, or Trustworthy AI) is not, as some may claim, a way to give machines some kind of 'responsibility' for their actions and decisions, and in the process discharge people and organisations of their responsibility*". *In:* DIGNUM, Virginia. Responsible Artificial Intelligence - from Principles to Practice. **ACM SIGIR Forum**, v. 56, n. 1, p. 4, jun. 2022. Disponível em: https://arxiv.org/pdf/2205.10785. Acesso em: 02 jun. 2024.

[64] BARBOSA, Mafalda Miranda. Responsabilidade Civil pelos danos causados por entes dotados de inteligência artificial. p. 170. *In:* BARBOSA, Mafalda Miranda; et al (org.). **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Editora Foco, 2021.

[65] Texto original: "*On the contrary, responsible development and use of AI requires more responsibility and more accountability from the people and organisations involved: for the decisions and actions of the AI applications, and for their own decision of using AI in a given application context*". *In:* DIGNUM, Virginia. Responsible Artificial Intelligence - from Principles to Practice. **ACM SIGIR Forum**, v. 56, n. 1, p. 4, jun. 2022. Disponível em: https://arxiv.org/pdf/2205.10785. Acesso em: 02 jun. 2024.

[66] Texto original: "**1.5. Accountability**: a) AI actors should be accountable for the proper functioning of AI systems and for the respect of the above principles, based on their roles, the context, and consistent with the state of the art. b) To this end, AI actors should ensure traceability, including in relation to datasets, processes and decisions made during the AI system lifecycle, to enable analysis of the AI system's outputs and responses to inquiry, appropriate to the context and consistent with the state of the art. c) AI actors, should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system

Apesar da Lei Geral de Proteção de Dados (LGPD) optar por não fazer recomendações especificas, nem vedar o tratamento de dados pessoais por meio de uso de algoritmos,[67] assim como não possui disposições expressas que mencionem os sistemas de Inteligência Artificial. A norma tangência a temática quando aborda, em seus artigos 49 e 50, § 1º, as boas práticas e a necessidade de adotar medidas de segurança ao utilizar sistemas para o tratamento de dados pessoais,[68] visando o *Privacy by Design*. Onde busca-se "a criação de confiança não apenas no método, no procedimento de coleta e de tratamento de dados, mas também na própria arquitetura do sistema".[69] Ou seja, também trabalha a responsabilidade por meio da prevenção.

> **Art. 49.** Os sistemas utilizados para o tratamento de dados pessoais devem ser estruturados de forma a atender aos requisitos de segurança, aos padrões de boas práticas e de governança e aos princípios gerais previstos nesta Lei e às demais normas regulamentares.
>
> **Art. 50.** (...).
> **§ 1º** Ao estabelecer regras de boas práticas, o controlador e o operador levarão em consideração, em relação ao tratamento e aos dados, a natureza, o escopo, a finalidade e a probabilidade e a gravidade dos riscos e dos benefícios decorrentes de tratamento de dados do titular [grifo nosso].[70]

Consequentemente, em decorrência da aplicação dos artigos mencionados acima, os princípios gerais previstos no artigo 6º da LGPD, também devem ser observados pelos sistemas de IA no Brasil, ao realizar o tratamento de dados pessoais. Destacando-se os seguintes princípios já aqui expostos e previstos em outros documentos:

> Art. 6º As atividades de tratamento de dados pessoais deverão observar a boa-fé e os seguintes princípios:
> (...).

---

lifecycle on an ongoing basis and adopt responsible business conduct to address risks related to AI systems, including, as appropriate, via co-operation between different AI actors, suppliers of AI knowledge and AI resources, AI system users, and other stakeholders. Risks include those related to harmful bias, human rights including safety, security, and privacy, as well as labour and intellectual property rights". *In:* OECD/LEGAL/0449. Recommendation of the Council on Artificial Intelligence. **OECD Legal Instruments**. 2024. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 20 jul. 2024.

[67] MENDES, Laura Schertel; MATTIUZO, Marcela; FUJIMOTO, Mônica Tiemy. Discriminação algorítmica à luza da Lei Geral de Proteção de Dados. p. 421-518. *In:* MENDES, Laura; DONEDA, Danilo; SARLET, Ingo Wolfgang *et al*. (Coord.) **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

[68] FERNANDES, Rafael Gonçalves; OLIVEIRA, Liziane Paixão Silva. A regulação do agir decisório disruptivo no judiciário brasileiro e a observância do princípio da precaução: juiz natural ou "juiz artificial"? **Revista Opinião Jurídica**, Fortaleza, v. 19, n. 30, p. 91- 117, jan./abr. 2021. Disponível em: https://periodicos.unichristus.edu.br/opiniaojuridica/article/view/3446. Acesso em: 05 mai. 2023.

[69] LEMOS, Ronaldo; BRANCO, Sérgio. Privacy by design: conceito, fundamentos e aplicabilidade na LGPD. p. 459 *In*: MENDES, Laura; DONEDA, Danilo; SARLET, Ingo Wolfgang *et al*. (Coord.) **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

[70] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 3 mai. 2024.

> VI - **transparência:** garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial;
> (...).
> X - **responsabilização e prestação de contas:** demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas [grifo nosso].[71]

Não obstante os benefícios que estes princípios possam trazer para a esfera jurídica. Há autores que compartilham um certo receio quanto a sua eficácia no caso concreto. E.g. referindo-se especificamente a algoritmos e ao princípio da transparência, é possível encontrar alguns óbices a sua suficiência. Ao ser utilizado em sistemas que possuem alguma aleatoriedade e/ou são capazes de se mudar (adaptar) ao longo do tempo a depender do contexto (e.g. sistemas de IA que utilizam modelos de *machine learning*); ou pela simples intangibilidade em alguns casos (e.g. razões públicas fundamentadas, como a segurança nacional, que afastam o direito à divulgação) e pela possível insuficiência de tornar transparente sistemas que por exemplo não foram projetados para considerar futuras avaliações e prestações de contas.[72]

"As normas éticas impõem condutas, as quais, se observadas, cumprem papel similar ao desempenhado pelas normas jurídicas. Contudo, não se pode deixar de reconhecer que as normas jurídicas são bem mais eficazes para a regulação das condutas humanas".[73] Apenas a implementação desses princípios durante o desenvolvimento e uso dos sistemas de IA pode não garantir uma proposta de política suficiente.[74] A tecnologia não é neutra e seu desenvolvimento deve se dar também em consonância com elas. Qualquer que seja a solução concreta a ser adotada, e.g. uma combinação dos diversos mecanismos aqui apresentados ou a sua aplicação de maneira isolada. A centralidade do papel humano no processo de automação sempre deve ser o objetivo principal a ser perquirido.[75]

---

[71] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 3 mai. 2024.
[72] KROLL, Joshua, A. *et al*. Accountable Algorithms. **University of Pennsylvania Law Review**, v. 165, 2017. Disponível em: https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/. Acesso em: 15 jun. 2024.
[73] FACHIN, Zulmar; FACHIN, Jéssica; SILVA, Deise Marcelino da. Princípios de inteligência artificial. **Constituição, Economia e Desenvolvimento: Revistada Academia Brasileira de Direito Constitucional**. Curitiba, 2022, vol. 14, n. 26, jan./jul.,2022. Disponível em: https://abdconstojs.com.br/index.php/revista/article/view/434/292. Acesso em: 06 jul 2024.
[74] MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **RDU**, Porto Alegre, Volume 16, n. 90, p. 20, nov-dez 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766/Schertel%20Mendes%3B%20Mattiuzzo%2C%202019. Acesso em: 06 jul. 2024.
[75] MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **RDU**, Porto Alegre, Volume 16, n. 90, p. 20, nov-dez 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766/Schertel%20Mendes%3B%20Mattiuzzo%2C%202019. Acesso em: 06 jul. 2024.

4. Considerações finais

Diante do exposto, verifica-se que a inteligência artificial, embora represente importante instrumento de inovação e desenvolvimento tecnológico, também apresenta riscos relevantes à proteção dos direitos fundamentais, especialmente no que se refere à proteção de dados pessoais e ao direito à igualdade. Os casos analisados demonstram que sistemas de IA não são neutros, podendo reproduzir e potencializar preconceitos existentes nos dados utilizados em seu treinamento ou nas escolhas realizadas por seus desenvolvedores. Nesse contexto, observa-se que o ordenamento jurídico brasileiro já dispõe de fundamentos legais relevantes para enfrentar parte desses desafios, ainda que persista a necessidade de maior amadurecimento regulatório específico sobre o tema.

Além disso, conclui-se que princípios de governança algorítmica, como transparência, explicabilidade, inteligibilidade e accountability, desempenham papel essencial na construção de sistemas de IA mais confiáveis e compatíveis com a tutela dos direitos fundamentais. Contudo, a mera adoção de diretrizes éticas não se mostra suficiente sem mecanismos efetivos de fiscalização, responsabilização e prevenção de danos. Assim, o desenvolvimento e a utilização da inteligência artificial devem ocorrer em consonância com os direitos fundamentais e com a preservação da centralidade humana no processo decisório, de modo que o avanço tecnológico não se transforme em instrumento de ampliação de desigualdades ou de esvaziamento das garantias constitucionais.

5. Referências

ACM Code of Ethics and Professional Conduct. **Association for Computing Machinery**. 2018. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 20 jul. 2024.

BARBOSA, Mafalda Miranda. Responsabilidade Civil pelos danos causados por entes dotados de inteligência artificial. p. 170. *In:* BARBOSA, Mafalda Miranda; et al (org.). **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Editora Foco, 2021.

BARROSO, Luís Roberto; OSÓRIO, Aline. "Sabe com quem está falando? ": Notas sobre o princípio da igualdade no Brasil contemporâneo. **Revista Direito e Práxis**, v. 7, n. 1, p. 207-208, 2016. Disponível em: https://www.e-publicacoes.uerj.br/index.php/revistaceaju/article/view/21094/15886. Acesso em: 01 abr. 2023.

BESSA, Leonardo Roscoe. A Lei Geral de Proteção de Dados e o direito à autodeterminação informativa. **Conjur**. 2020. Disponível em: https://www.conjur.com.br/2020-out-26/leonardo-bessa-lgpd-direito-autodeterminacao-informativa.. Acesso em: 07 de ago de 2023.

BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019, p. 94-132. *E-book.*

BOBBIO, Norberto. **Democracia e Segredo**. São Paulo: UNESP, 2015. p. 33.

BRASIL, Ministério da Ciência, Tecnologia e Inovações. **Inteligência Artificial**. Brasília, 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial. Acesso em: 20 de ago. de 2023.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 mai. 2024.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Brasília, DF: Presidência da República, [2023]. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 3 mai. 2024.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, [1990]. Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 20 out. 2024.

BRASIL. **Mensagem nº 288, de 8 de julho de 2019. Brasília: Presidência da República, 2019. Disponível em: https://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Msg/VEP/VEP-288.htm. Acesso em: 17 abr. 2023.**

CALO, Ryan. Robotics and the lessons of cyberlaw. **California Law Review**, Los Angeles, v. 103, p. 542, jun. 2015. Disponível em: https://digitalcommons.law.uw.edu/faculty-articles/23/. Acesso em: 23 mai. 2024.

COPPIN, Ben. **Inteligência Artificial**. Rio de Janeiro, 2010. p. 3.

Declaração de Montreal pelo Desenvolvimento Responsável da Inteligência Artificial 2018. **Declaração de Montreal pela IA responsável**. 2018. Disponível em: https://declarationmontreal-iaresponsable.com/. Acesso em: 20 jul. 2024.

DIAKOPOULOS, Nicholas *et al*. Principles for Accountable Algorithms and a Social Impact Statement for Algorithms. **FAT/ML**. Disponível em: https://www.fatml.org/resources/principles-for-accountable-algorithms. Acesso em: 20 jul. 2024.

DOSHI-VELEZ, Finale *et al*. Accountability of AI Under the Law: The Role of Explanation. **Harvard Public Law**, 2017. Disponível em: https://www.researchgate.net/publication/320867322_Accountability_of_AI_Under_the_Law_The_Role_of_Explanation. Acesso em 06 jul. 2024.

Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems. **IEEE Advancing Technology for Humanity**. 2019. Disponível em: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf. Acesso em: 20 jul. 2024.

EUROPEAN Group on Ethics in Science and New Technologies. Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. **European Commission**. mar. 2018. Disponível em: https://www.unapcict.org/sites/default/files/2019-01/EC_AI-%20Robotics-%20and%20Autonomous%20Systems.pdf. Acesso em: 20 jul. 2024.

FACHIN, Zulmar; FACHIN, Jéssica; SILVA, Deise Marcelino da. Princípios de inteligência artificial. **Constituição, Economia e Desenvolvimento: Revistada Academia Brasileira de Direito Constitucional**. Curitiba, 2022, vol. 14, n. 26, p. 371, jan./jul.,2022. Disponível em: https://abdconstojs.com.br/index.php/revista/article/view/434/292. Acesso em: 06 jul 2024.

FERNANDES, Rafael Gonçalves; OLIVEIRA, Liziane Paixão Silva. A regulação do agir decisório disruptivo no judiciário brasileiro e a observância do princípio da precaução: juiz natural ou "juiz artificial"? **Revista Opinião Jurídica**, Fortaleza, v. 19, n. 30, jan./abr. 2021. Disponível em: https://periodicos.unichristus.edu.br/opiniaojuridica/article/view/3446. Acesso em: 05 mai. 2023.

FUTURE OF LIFE INSTITUTE. Policymaking in the Pause What can policymakers do now to combat risks from advanced AI systems? **Future of Life Institute**. 2023. Disponível em: https://futureoflife.org/wp-content/uploads/2023/04/FLI_Policymaking_In_The_Pause.pdf. Acesso em: 23 de ago. de 2023.

GODOY, Claudio Luiz Bueno de. A intervenção humana no procedimento de explicação e de revisão de decisões automatizadas. p. 287-309. *In:* CHINELLATO, Silmara J. de A.; TOMASEVICIUS FILHO, Eduardo. **Inteligência Artificial**: Visões interdisciplinares e internacionais. São Paulo: Almedina, 2023.

HARDESTY, Larry. Study finds gender and skin-type bias in commercial artificial-intelligence systems. **MIT News**. 2018. Disponível em: https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212. Acesso em: 22 out. 2024.

High-Level Expert Group on Artificial Intelligence (AI HLEG). Ethics guidelines for trustworthy AI. **European Commission**. Abr. 2019. Disponível em: https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf. Acesso em: 20 jul. 2024.

HOUSE of Lords. AI in the UK: ready, willing and able? **Select Committee on Artificial Intelligence** abr. 2019. Disponível em: https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf. Acesso em: 20 jul. 2024.

KROLL, Joshua, A. *et al*. Accountable Algorithms. **University of Pennsylvania Law Review**, v. 165, 2017. Disponível em: https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3/. Acesso em: 15 jun. 2024.

LEE, Peter. Learning from Tay's introduction. **Official Microsoft Blog**. 2016. Disponível em: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/. Acesso em: 22 out. 2024.

LEMOS, Ronaldo; BRANCO, Sérgio. Privacy by design: conceito, fundamentos e aplicabilidade na LGPD. p. 459 *In*: MENDES, Laura; DONEDA, Danilo; SARLET, Ingo Wolfgang *et al*. (Coord.) **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

LIMA, Cíntia Rosa Pereira de; OLIVEIRA, Cristina Godoy Bernardo de; RUIZ, Evandro Eduardo Seron. Inteligência artificial e personalidade jurídica: aspectos controvertidos. *In:* BARBOSA, Mafalda Miranda; et al (org.). **Direito digital e inteligência artificial: diálogos entre Brasil e Europa**. Indaiatuba, SP: Editora Foco, 2021.

MATTIUZZO, Marcela. Discriminação algorítmica: reflexões no contexto da Lei Geral de Proteção de Dados Pessoais, p.118. *In:* MENDES, Laura; DONEDA, Danilo; CUEVA, Ricardo Villas Bôas. **Lei Geral de Proteção de Dados (Lei 13.709/2018): a caminho da efetividade**. São Paulo: Revista dos Tribunais, 2020.

MCCARTHY, John; *et al*. **A proposal for the Dartmouth summer research project on artificial intelligence**. 1955. Disponível em: http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf. Acesso em: 20 jun. 2023.

MENDES, Laura Schertel; MATTIUZO, Marcela; FUJIMOTO, Mônica Tiemy. Discriminação algorítmica à luza da Lei Geral de Proteção de Dados. p. 421-518. *In:* MENDES, Laura; DONEDA, Danilo; SARLET, Ingo Wolfgang *et al*. (Coord.) **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

OECD AI Principles overview. **OECD.AI Policy Observatory**. 2024. Disponível em: https://oecd.ai/en/ai-principles. Acesso em: 20 jul. 2024.

OECD/LEGAL/0449. Recommendation of the Council on Artificial Intelligence. **OECD Legal Instruments**. 2024. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 20 jul. 2024.

REQUIÃO, Maurício; COSTA, Diego. Discriminação algorítmica: ações afirmativas como estratégia de combate. **civilistica.com**, [S. l.], v. 11, n. 3, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 15 mar. 2024.

SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. **Revista USP**. 2021. Disponível em: <https://www.revistas.usp.br/eav/article/view/185024/171209>. Acesso em: 06 jul. 2024.

SILVA, Priscila; MEDEIROS, Juliana. A polêmica da revisão (humana) sobre decisões automatizada. ITS Rio, Rio de Janeiro, 2019. Disponível em: https://feed.itsrio.org/a-pol%C3%AAmica-da-revis%C3%A3o-humana-sobre-decis%C3%B5es-automatizadas-a81592886345. Acesso em: 17 abr. 2023.

SIQUEIRA, Dirceu Pereira; MORAIS, Fausto Santos de; TENA, Lucimara Plaza. O papel emacipador do direito em um contexto de linhas abissais e algoritmos. **Pensar: Revista de Ciências Jurídicas**, v. 7, n. 1, p. 9, 2022. Disponível em: https://ojs.unifor.br/rpen/article/view/12058/6780. Acesso em: 27 set. 2024.

THE ASIMOLAR AI Principles. **The future of life**. 2017. Disponível em: https://futureoflife.org/open-letter/ai-principles/. Acesso em: 20 jul. 2024.

TURING, Alan M. Computing Machinery and Intelligence. **Oxford University Press on behalf of the Mind Association**, New Series, v. 59, n. 236, out. 1950), p. 433-460. Disponível em: https://phil415.pbworks.com/f/TuringComputing.pdf. Acesso em: 23 jun. 2023.

WARREN, Samuel D.; BRANDEIS, Louis D. The right to privacy**. Harvard Law Review**, Vol. 4, n. 5, 1890, p. 193-220. Disponível em: https://www.jstor.org/stable/1321160. Acesso em: 07 de ago de 2023.