Uma Análise Jurídica e Sociológica dos Vieses Algorítmicos na Inteligência Artificial sob a Perspectiva da Ética dos Dados

An Analysis of Algorithmic Biases in Artificial Intelligence from a Legal and Sociological Perspective of Data Ethics

Adriel Mafra Limas - Universidade do Vale do Itajaí (UNIVALI), adrielmafra05@gmail.com
Arthur Ogliari Lana - Universidade do Vale do Itajaí (UNIVALI), ogliari.arthur@gmail.com
José Everton da Silva - Universidade do Vale do Itajaí (UNIVALI), caminha@univali.br

Resumo:
O presente artigo busca identificar os problemas decorrentes do uso desregulado da inteligência artificial, especialmente a controvérsia em torno dos vieses algorítmicos, bem como apontar a ética como possível solução sociológica e jurídica para salvaguardar direitos e interesses individuais e coletivos potencialmente afetados por sistemas inteligentes. O objetivo geral é analisar a (in)existência de vieses algorítmicos em sistemas autônomos dotados de inteligência artificial. Como objetivos específicos, pretendem-se: identificar como vieses humanos são implementados, intencionalmente ou não, em máquinas; examinar de que modo a ética pode auxiliar, sociologicamente, na proteção dos direitos humanos frente ao uso de sistemas inteligentes pela iniciativa pública e privada; verificar se o Direito pode admitir a ética como instrumento de tutela desses direitos; e analisar casos concretos. O problema de pesquisa consiste em compreender como a ética pode contribuir, sociologicamente e juridicamente, para coibir vieses discriminatórios praticados por sistemas autônomos. Ao final, verificou-se que a ética dos dados pode servir, em termos sociológicos, como instrumento assecuratório de direitos humanos fundamentais, havendo amplo debate social e acadêmico acerca dos riscos do uso desregulado da inteligência artificial. No âmbito jurídico, especialmente na atividade decisória dos magistrados, constatou-se divergência doutrinária quanto à possibilidade de utilização da ética dos dados na solução de casos concretos envolvendo violações de direitos decorrentes de máquinas automatizadas. Quanto à metodologia, adotou-se o método indutivo, com fontes primárias e secundárias.
Palavras-chave: Ética dos dados; inteligência artificial; vieses cognitivos.

Abstract:
This article seeks to identify the problems arising from the unregulated use of artificial intelligence, especially the controversy surrounding algorithmic bias, and to present ethics as a possible sociological and legal solution to safeguard individual and collective rights and interests that may be affected by intelligent systems. The general objective is to analyze the (non)existence of algorithmic bias in autonomous systems equipped with artificial intelligence. The specific objectives are: to identify how human biases are intentionally or unintentionally embedded in machines; to examine how ethics may sociologically assist in protecting human rights against the use of intelligent systems by both public and private entities; to verify whether the law may recognize ethics as an instrument for safeguarding such rights; and to analyze concrete cases. The research problem asks how ethics, from both sociological and legal perspectives, may help restrain discriminatory biases in autonomous systems. The study found that data ethics may, sociologically, serve as a tool for protecting human and fundamental rights, while social and academic debate continues over the risks of unregulated artificial intelligence. In the

legal sphere, particularly in judicial decision-making, doctrinal disagreement arose over the use of data ethics in resolving cases involving rights violations caused by automated systems. As for methodology, the study employed an inductive approach.
Keywords: Data ethics; artificial intelligence; algorithmic biases.

1. Introdução

Com o desiderato de garantir que o desenvolvimento tecnológico de algoritmos1 caminhe lado a lado às normas de direitos humanos e valores democráticos, em maio de 2019 foi concebido pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) o documento intitulado "Recomendation of the Council on Artificial Intelligence" (Recomendação do Conselho sobre Inteligência Artificial), o qual traça diretrizes que – acaso implementadas – podem estabelecer um padrão em escala intergovernamental no tocante ao acertado uso da inteligência artificial2 (IA).

Pode-se afirmar que o aludido documento internacional surge como corolário direto de uma demanda que, cada vez mais, atinge a população mundial: a premente necessidade de delinear diretrizes éticas e maior accountability na implementação de novas tecnologias pelos governos.

No ponto, ressalta-se a salutar relevância da Recomendação, ao passo que "os 36 países membros da OCDE, juntamente com Argentina, Brasil, Colômbia, Costa Rica, Peru e Romênia, assinaram os Princípios da OCDE sobre Inteligência Artificial na reunião anual do Conselho Ministerial da Organização" (CORTELINI et al, 2020, p. 425-426).

Sob essa perspectiva, merece menção o documento "G20 AI Principles", que – projetando as recomendações da OCDE e destacando a importância da inovação e confiança em tecnologias digitais – traçou diversos princípios a serem seguidos em âmbito internacional quando da programação de sistemas computacionais inteligentes, a saber: a promoção do crescimento tecnológico inclusivo, desenvolvimento sustentável e bem-estar social; proteção aos valores de justiça centrados na pessoa humana;

[1] **Nota explicativa:** Adota-se, em suma, a conceituação de algoritmo apresentada por Franco Pizzetti, segundo a qual o algoritmo é entendido como uma expressão matemática que traduz um procedimento lógico e sistemático de cálculo, servindo de base para o funcionamento dos sistemas de inteligência artificial.

[2] **Nota explicativa:** Seguindo a definição utilizada pelo Nicolò Basigli em sua tese *(La possibilità della repubblicanizzazione dell'intelligenza artificiale, 2022)*, a inteligência artificial representa algo distinto da mera automação de processos existentes: originariamente baseada na criação de algoritmos em condições de dar respostas coerentes a inputs correspondentes (Turing), hoje consiste na programação de sistemas de análise que se evoluem e 'melhoram' com base na própria experiência (*machine learning*), segundo a intuição de Arthur L. Samuel (1959) e posteriormente estruturada por Tom Mitchell (1997).

transparência e explicabilidade; robustez, segurança, proteção e; *accountability* (OCDE, 2019, p. 11-12).

De forma derradeira, em novembro de 2021 a UNESCO aprovou a Recomendação sobre a Ética da Inteligência Artificial, primeiro instrumento normativo global de caráter universal, que estabelece princípios para garantir que a IA seja desenvolvida e utilizada em conformidade com os direitos humanos, a dignidade e a justiça social (UNESCO, 2021)

Por fim, em meados de 2024, o Parlamento Europeu aprovou o *Artificial Intelligence Act,* considerado o primeiro marco regulatório abrangente sobre a matéria, que classifica sistemas de IA em níveis de risco e impõe obrigações específicas a fornecedores e usuários, tornando-se referência normativa internacional (EU, Regulamento 1689, 2024).

Ora, não obstante a inclinação do ser humano em crer que as tecnologias são essencialmente neutras, é notório que muitos valores negativos podem – mesmo que não intencionalmente – ser inseridos nos algoritmos de uma máquina. Daí surge o conceito de "vieses discriminatórios", que acabam por macular a IA e, consequentemente, ferir princípios éticos e morais positivados em diplomas internacionais, ou mesmo na própria Constituição dos países.

Nessa vereda, ensina Matheus Oliveira Fornasier:

> Muitos sistemas de aprendizado de máquina utilizam grandes quantidades de dados (coletados por meio de mídias sociais, aplicativos, telefones celulares, cartões de crédito etc.) sobre o comportamento humano e, em relação a isso, pesquisadores já descobriram práticas discriminatórias atinentes a esses sistemas (na justiça criminal, na pontuação de crédito e na publicidade, por exemplo). Assim, o aprendizado de máquina justo é um novo campo de estudo para mitigar vieses discriminatórios incorporados aos algoritmos (FORNASIER, 2021, p. 177).

Dessa forma, infere-se que coibir tais vieses discriminatórios embutidos na programação algorítmica de máquinas – sejam elas derivadas da iniciativa privada ou pública – é medida necessária para garantir o respeito aos direitos dos usuários de programações inteligentes.

Nessa conjuntura, Pimentel e Orengo fazem referência ao surgimento de um novo campo da ética denominado "ética dos dados" ou "*data ethics*", que visa justamente assegurar princípios como os abalizados pela OCDE:

> [...] não se pode contestar que a ética, como ramo do conhecimento humano, também evolui, de modo que é possível falar-se na

> contemporaneidade em um campo de ética dos dados utilizados pelas novas tecnologias (*data ethics*), o qual se preocupa com a análise de relevantes temas éticos como privacidade, anonimato, transparência, confiança e responsabilidade, com concentração em seus aspectos procedimentais (PIMENTEL; ORENGO, 2021, p. 319).

Assim, pode-se afirmar que a ética dos dados, orientada pelas recomendações internacionais e pelos princípios constitucionalmente estabelecidos, tem por pretenso promover um caráter universalista de bons princípios, com o "compartilhamento no âmbito mundial de determinados conjuntos de valores, de modo a otimizar a relação harmônica entre todos os povos da terra" (ZANON JUNIOR, 2014, p. 15).

Nessa toada, na concepção de Orlando Luiz Zanon Júnior, o Direito surge como uma especialização da ética, a saber: uma positivação *a posteriori* dos valores criados por determinada sociedade, prevendo, ademais, as consequências do descumprimento das condutas moralmente aceitáveis por determinado conjunto populacional, sem prejuízo de sobreposição ou ampliação do ordenamento jurídico com a possibilidade de intervenção de órgãos supranacionais. Veja-se:

> Elaborando a partir daí, pode-se afirmar que o Direito é, em verdade, uma especialização da Ética, no sentido de fixar no nível estatal os padrões decisórios a serem observados em sociedade, bem como estabelecer as respectivas consequências, sejam elas promocionais, reparativas, punitivas ou quaisquer outras, a serem reforçadas por estruturas organizacionais especialmente designadas para tal finalidade [...]. Outrossim, o Direito é um passo adiante da Ética, no sentido da institucionalização de padrões de conduta mediante a articulação da Sociedade com o Estado, sem que nada veda eventual sobreposição ou ampliação do sistema, mediante a criação de órgãos supranacionais e internacionais, sem prejuízo de influxos decorrentes do pluralismo normativo ou da transnacionalidade.

Diante disso, propõe-se a presente pesquisa a investigar a temática dos vieses algorítmicos e, por conseguinte, a analisar a ética de dados como instrumento de repressão às práticas discriminatórias a eles associadas.

Isto posto, assume-se como problema de pesquisa o seguinte questionamento: Como a ética pode auxiliar, sociologicamente e juridicamente, na coibição de vieses discriminatórios perpetrados por sistemas autônomos?

Por outro lado, tem-se como objetivo geral do artigo analisar a (in)existência de vieses algorítmicos em sistemas autônomos dotados de Inteligência Artificial. Nesse sentido, os objetivos específicos são: 1) identificar como os vieses humanos são,

intencionalmente ou não, implementados em máquinas; 2) analisar como a ética pode auxiliar, sociologicamente, na salvaguarda dos direitos humanos frente ao uso de máquinas dotadas de inteligência artificial programadas pela iniciativa pública e privada; 3) analisar se o direito pode admitir a ética como instrumento de salvaguarda aos direitos humanos frente ao uso de máquinas dotadas de inteligência artificial e; 4) verificar casos concretos.

Consigna-se, que as definições de nomenclaturas técnicas como "inteligência artificial", "algoritmos", e outras acepções correlatas não são unívocas e exigem um exame interdisciplinar da matéria. Destarte, a fim de dar prosseguimento aos objetivos do presente artigo, dispensar-se-á o aprofundado tratamento terminológico desses termos.

Quanto à metodologia científica adotada nesta pesquisa, utilizou-se como parâmetro o método indutivo. Também, foram aproveitadas como técnicas auxiliares à metodologia de pesquisa, a técnica do fichamento, o referente, a categoria e o conceito operacional.

2 Da inserção de vieses algorítmicos em sistemas automatizados

Em julho de 2016, um concurso de beleza patrocinado pela empresa Beauty AI[3] despertou o clamor midiático ao anunciar – de forma inusitada – que os julgamentos seriam realizados por um júri composto por robôs, que analisariam critérios objetivos (rugas, simetria, medidas faciais etc.) e, por fim, anunciariam os vencedores do campeonato. A proposta visava justamente afastar preconceitos no padrão decisório dos julgadores, mas o resultado foi expressivamente contrário ao esperado (GARCIA, 2020, p. 16-17).

Mais de 6.000 (seis mil) pessoas haviam enviado fotografias de seus rostos para participar da disputa, mas, dos 44 (quarenta e quatro) vencedores, apenas um tinha pele escura, em que pese o número de indivíduos de diferentes raças inscritos ser proporcional (GARCIA, 2020, p. 17).

Diante da polêmica, os desenvolvedores demonstraram que o sistema não havia sido programado para avaliar peles claras como um sinal adicional de beleza. Todavia, comprovou-se que a base de dados dos robôs/julgadores havia sido alimentada com

[3] Mais informações em: https://beauty.ai . Acesso em 30 ago. 2025.

diversas imagens de atores e atrizes hollywoodianos, que, em sua maioria, eram brancos, o que influenciou diretamente o resultado do concurso (GARCIA, 2020, p. 17).

Por óbvio, o uso problemático de máquinas enviesadas por governanças públicas tende a trazer um abalo muito maior à população. É o que se pode inferir da desastrosa experiência dos Estados Unidos da América, o que demonstra que nem mesmo a maior potência mundial ficou livre de polêmicas envolvendo o uso da inteligência artificial:

> Nos Estados Unidos [...] vinte e quatro estados aderiram a algoritmos que prometem um sistema mais imparcial e livre de vieses, para tornar as sentenças judiciais mais homogêneas e previsíveis. [...] O mais famoso desses modelos, o LSI-R, inventado em 1995, consiste em um longo questionário que os condenados devem responder e que será utilizado para "prever" a probabilidade de reincidência, classificando esse risco em alto, médio ou baixo. Dentre as perguntas que devem ser respondidas pelos réus figuram: "quantas condenações anteriores você teve?" ou "qual a influência de drogas ou álcool na prática do crime?" O questionário também traz questões relativas a circunstâncias de nascimento e formação do condenado, inclusive perguntas sobre o histórico criminal de membros de sua família e amigos. Mesmo com o fator aleatório decorrente de decisões humanas diminuído, alguém pode questionar se os vieses foram eliminados ou apenas camuflados. Ainda que não contenha perguntas que digam respeito expressamente à etnia, não é difícil imaginar como um modelo que leva em conta tais variáveis apresentará um viés racial. É evidente que pessoas vindas de bairros de baixa renda (mais policiados) terão maior probabilidade de já terem tido prévios contatos com a polícia, se comparados com pessoas de classe média ou alta em face do processo de seleção secundária denunciados pela Criminologia Crítica. Igualmente, será mais provável que alguém de sua família tenha algum histórico criminal ou problemas relacionados ao álcool e às drogas. Nesse cenário, o modelo sequer precisa analisar uma resposta relativa à raça para se tornar racista, pois tal informação é desnecessária quando as demais já apontam para questões racialmente sensíveis. O risco de correlações ilusórias enviesa os resultados (BOEING; ROSA, 2020, p. 86).

Assim como os desenvolvedores da empresa Beauty AI e o governo estadunidense, muitas pessoas tendem a depositar extrema confiança em decisões advindas de algoritmos, ao passo que acreditam veementemente que estes promoverão resoluções mais precisas e livres de preconceitos, de informações inverídicas ou sem fundamentos (ALENCAR, 2021, p. 77).

Entretanto, essa linha de raciocínio se revela, em última análise, distanciada da realidade, consoante ensina Ana Catarina Alencar:

> Na contramão do sistema de confiança esperado, a tecnologia tem se mostrado exclusiva em muitos momentos. Isso ocorre, especialmente, quando algoritmos de redes sociais utilizadas no mundo todo criam “bolhas” de entretenimento e informação personalizada, gerando

> polarização social e política, além de propiciar a disseminação de informações falsas e discurso de ódio nas redes (ALENCAR, 2021, p.81. Grifou-se).

Para entender como uma máquina dotada de inteligência artificial pode adotar práticas discriminatórias, é imprescindível compreender como as novas tecnologias podem "aprender".

Daniel Henrique Arruda Boeing e Alexandre de Morais da Rosa explicam o aprendizado de máquina (*machine learning*) – exclusivo de inteligências artificiais fortes, de propósito geral ou de nível humano[4] – da seguinte forma:

> O processo de "aprendizado", em uma rede neural artificial, ocorre através da manipulação de "pesos e vieses". Cada neurônio possui um viés (valor mínimo) de ativação e está conectado a todos os neurônios da camada seguinte. Para ser ativado, e com isso propagar informações para o próximo "neurônio", o impulso (*input*) recebido deve ser suficientemente forte (estar acima de um determinado valor) para romper tal limiar, de forma que ele gera um *output*, que servirá para ativar (ou não) o próximo neurônio da rede, processo que se dá em cascata. Além dos vieses, cada uma dessas conexões possui um "peso". Dessa forma um impulso mais fraco pode ser potencializado por uma conexão de maior força (e assim ativar o próximo neurônio), ao passo que um output mais forte pode ser mitigado se sua conexão tiver um peso menor. "Aprender", no caso, significa otimizar sua performance, sem intervenção humana, manipulando tais valores numéricos, o que servirá para fazer predições em novos conjuntos de dados (BOEING; ROSA, 2020, p. 26).

Por seu turno, Hugo Brito Machado Segundo – além de realçar como os vieses podem ser (in)voluntariamente inseridos em programas – ilustra, de forma simples e elucidativa, a dinâmica da programação dos dados:

> A partir de informações com as quais são alimentados (*input*), são capazes de executar as fórmulas matemáticas que os integram e, em seguida, apresentar resultados (*output*). Nessa condição, se os dados dos quais partirem forem enviesados, os resultados possivelmente também o serão. Carl T Bergstrom e Jevin West, no curso *Calling Bullshit in the Age of Big Data*, disponível no Youtube, usam expressões simples e diretas para designar esse fenômeno: *garbage in*, *garbage out*. O que pode ser traduzido como: se entra lixo, o que sai também é lixo. Voltando ao exemplo da receita de bolo, para ilustrar, de modo muito simplificado, o que seria um algoritmo, ainda que se tenha uma excelente receita, o

[4] **Nota explicativa:** Segundo John Searle (1980), a chamada IA fraca corresponde a sistemas especializados em tarefas específicas, sem consciência ou compreensão, enquanto a IA forte se refere, em termos hipotéticos, a máquinas dotadas de inteligência de propósito geral, comparável à humana. No mesmo sentido, Nick Bostrom (2015) utiliza a expressão "superinteligência" para designar sistemas que, caso venham a existir, poderiam superar a inteligência humana em praticamente todos os domínios.

> emprego de ingredientes de má qualidade prejudicará, sem dúvida, o resultado (MACHADO SEGUNDO, 2023, s. p.).

Com efeito, Flávio Iassuo Takakura e Luciana Gaspar Melquíades Duarte desvendam que o aprendizado de máquina (*machine learning*) é subgrupo da inteligência artificial, apontando, por conseguinte, suas classificações:

> A IA é o grande grupo de sistemas computacionais que se propõe a simular a inteligência humana e que contém um subgrupo, o *machine learning (*aprendizado de máquina, em que se encontram os algoritmos que executam essa tarefa de pensar. Esse subgrupo pode ser classificado como supervisionado, não supervisionado ou de aprendizado por reforço (TAKAKURA; DUARTE, 2022, p. 5).

Pois bem. No aprendizado de máquina supervisionado tem-se, necessariamente, o acompanhamento de um ser humano que fornece dados à máquina, gerando a dinâmica de *input* e *output* e a consequente previsibilidade das ações tomadas pelos algoritmos. Doutra forma, no aprendizado não supervisionado, a própria IA classifica os dados, enquanto, no aprendizado por reforço, a máquina aprende por tentativa e erro, hipóteses que dão ainda mais vazão ao desenvolvimento de vieses pelo próprio autômato.

Não bastasse, ao inserir nessa dinâmica programática uma inteligência artificial dotada de deep learning (aprendizado profundo), seria possível afirmar que a imprevisibilidade dos resultados alcançados pela própria máquina se intensificaria, diferentemente do que acontece no machine learning, em que a máquina define resultados intrinsecamente transparentes, ao passo que, sendo eles discriminatórios, poderiam ser elididos pelo ser humano que a programa.

São nesse sentido os ensinamentos de Flávio Iassuo Takakura e Luciana Gaspar Melquíades Duarte:

> A existência de mais camadas implica uma grande alteração em relação ao modelo de uma única camada. No de uma camada, *machine learning*, dado um conjunto de dados, é possível inferir quais deram origem ao resultado, ou seja, o algoritmo é transparente – se não para o público em geral, ao menos para os especialistas. Já no caso do *deep learning*, isso não ocorre. O algoritmo é opaco, pois, com a introdução de mais camadas de "neurônios", quanto mais camadas se tem, mais opaco se torna, no sentido de que, dado um resultado, não é, em tese, possível saber quais caminhos foram percorridos pelos dados que deram origem ao mesmo (TAKAKURA; DUARTE, 2022, p. 6).

Concluem os referidos autores que a inserção de vieses pode ocorrer pelo *software* adotado pelo programador ou pelos dados iniciais enviesados por ele utilizados. Em

ambos os casos, "faz-se necessária uma intervenção externa para correção" (TAKAKURA; DUARTE, 2022, p. 6-7).

Assim sendo, considerando que o viés do algoritmo pressupõe um viés do próprio programador, percebe-se que necessariamente os vieses dos algoritmos estão atrelados à pré-existência de vieses cognitivos de seres humanos, beirando o impossível de se falar – no atual cenário – em inteligências artificiais fortes livres de vícios, o que agrava de sobremaneira a possibilidade de transgressão aos direitos da população mundial quando do uso de novas tecnologias.

Insta salientar que a psicologia cognitiva tem demonstrado a existência de uma infinidade de vieses cognitivos, que podem, por via de consequência, ser inseridos em sistemas autônomos programados por seres humanos susceptíveis a falhas.

Nesse prisma temático, colhem-se as lições de George Marmelstein:

> Há vieses decorrentes da limitação da mente humana em processar todas as informações relevantes, gerando falhas de atenção, percepção e julgamento. Outros que derivam de uma busca por conforto cognitivo, alimentados por atitudes, emoções e crenças que orientam o pensamento em direção à confirmação das expectativas. Outros decorrem de um processo de categorização equivocada e de associações implícitas e automáticas, que levam a generalizações apressadas e infundadas. Outros fatores de influência social que estimulam um comportamento de rebanho (conformidade e obediência) e um desejo de aceitação, e assim por diante (MARMELSTEIN, 2022, p. 233).

Para entender melhor a dimensão do assunto, John Manoogian III catalogou 188 (cento e oitenta e oito) tipos de vieses cognitivos em um *Cognitive Bias Codex* (códice de viés cognitivo)[5] (MARMELSTEIN, 2022, p. 233).

Em apertada síntese, David Danks e Alex John London propõem 5 (cinco) categorias de vieses algorítmicos, nos moldes delineados:

> **a)** Viés nos dados de treinamento (*Training Data Bias*): o desvio está localizado nos dados de treinamento ou na entrada (input). Esse tipo de viés será difícil de identificar, uma vez que as companhias raramente divulgam os dados utilizados na alimentação de um sistema. Nesses casos, usualmente, os dados de saída do sistema (*outputs*) refletirão vieses sociais mais amplos (*broader social biases*).
> **b)** Viés decorrente de algoritmo com foco seletivo de dados (Algorithmic Focus Bias): esse tipo de desvio ocorre quando "há um uso diferencial da informação contida no input ou dados de treinamento". Em outras palavras, esse viés surge quando a IA deliberadamente opta por não utilizar parte da informação, o que leva a um resultado enviesado.

[5] A íntegra da imagem-resumo elaborada por John Manoogian III pode ser visualidade em http://tinyurl.com/thebiascodex. Acesso em 30 ago. 2025.

> **c)** Viés algorítmico de processamento (*Algorithmic Processing Bias*): quando há uma escolha deliberada de inserir desvios no algoritmo com propósitos futuros, tais como no sentido de mitigar eventuais vieses que se sabem presentes nos dados de treinamento ou entrada.
> **d)** viés de transferência de contexto (*Transfer Context Bias*): esse tipo de desvio decorre do uso indevido do algoritmo, ou seja, emerge da aplicação ou extensão imprópria do algoritmo para além do contexto para o qual ele foi concebido". Não se pode replicar o modelo indiscriminadamente.
> **e)** Viés de interpretação (*Interpretation Bias*): ocorre quando há uma interpretação equivocada dos dados de saída (*outputs*) ou do funcionamento do algoritmo, tanto pelo usuário quanto pelo sistema autônomo no qual o algoritmo opera. Em outras palavras, no viés de interpretação, mesmo quando opera no contexto pretendido, observa-se uma discrepância entre "(i) a informação que o algoritmo produz; e (ii) os requisitos de informação fornecidos pelo usuário ou pelo sistema que usa o *output*". A fim de possibilitar uma melhor compreensão desse tipo de viés, convidamos o(a) leitor(a) a pensar em um sistema de monitoramento autônomo voltado à identificação de indivíduos que requerem vigilância (*surveillance value* de diferentes indivíduos). Em face das incontáveis possíveis interpretações ou conteúdo semântico da especificação "*surveillance value*", abre-se a possibilidade para a interpretação equivocada dessa especificação e a consequente manifestação de vieses no *output*.

Nesse cenário, evidencia-se a precariedade das normas reguladoras de mecanismos programados com inteligência artificial, evidenciando o risco para a população diante das novas tecnologias. É justamente por isso que a ética demonstra, em primeira análise, ser um importante mecanismo qualificado para enfrentar os vieses de máquina, tanto sociológicos como jurídicos, nos moldes delineados no tópico sequente.

3 A ética como solução sociológica e jurídica às discriminações algorítmicas

3.1 Um breve ensaio terminológico a respeito da ética e da moral

Antes de adentrar no mote principal da presente pesquisa, vale esclarecer a linha tênue que divide os institutos da “ética" e da “moral”, valendo-se, para tanto, das lições de Orlando Luiz Zanon Júnior, o qual afirma que a ética consiste em "um conjunto de parâmetros valorativos convergentes em determinado grupo de pessoas ou comunidade individualmente identificável, de modo a orientar e delimitar as deliberações tomadas perante o agrupamento” (ZANON JUNIOR, 2014, p. 24), enquanto a moral se traduz como “a escala de valores de cada pessoa, voltada ao direcionamento daquilo que é certo ou

errado (justo ou injusto), de acordo com seu conhecimento adquirido, de modo a orientar as suas deliberações" (ZANON JUNIOR, 2014, p. 13).

Nessa linha, o autor esclarece, de forma sintetizada, a diferença entre as acepções:

> [...] os dois institutos dizem respeito ao direcionamento do que deve ou não deve ser feito, de acordo com a aferição axiológica daquilo que é correto e incorreto (justo ou injusto), com vistas a limitar a capacidade de deliberação, na medida em que tendem a restringir certas condutas ao classificá-las como negativas, indignas, indevidas ou inadequadas. Porém, enquanto a Moral revela uma apreciação individual acerca do que é certo ou errado, a Ética representa a convergência desses parâmetros de correção para um determinado grupo, maior ou menor, mediante um diálogo consensual construído ao longo do tempo, ou mesmo por meio da imposição de alguma força social prevalecente no respectivo agrupamento. (ZANON JÚNIOR, 2014, p. 15. Grifou-se).

Doutra forma, alguns doutrinadores explicam a ética como gênero, que comporta três espécies: i) normas de etiqueta, que disciplinam atos de decoro, sendo que sua transgressão resulta em ato de mera descortesia, gerando sanção social de índole difusa (isto é, não institucionalizada, que pode ser exercida por todos membros do corpo social), por meio de manifestações concretas de ostracismo; ii) normas morais, que disciplinam questões éticas mais relevantes para garantia do harmonioso convívio social, cujo descumprimento resulta em um ato de imoralidade, forma mais grave de sanção social, que ocasiona a sanção difusa; iii) normas jurídicas, que estabelecem regras indispensáveis de sobrevivência ao corpo social, sendo que seu descumprimento gera ilicitude, sucedida pela sanção do próprio Estado, que detêm o monopólio para aplicação da sanção jurídica (SOARES, 2019, p. 13-17).

Não obstante, a visão adotada no presente estudo será a inaugural, isto é, a apresentada pelo professor Zanon Filho, segundo a qual a ética provém da construção social de determinados valores, enquanto a moral é a adoção individualizada desses valores.

Feito este breve prefácio, passa-se à análise de como a ética pode impactar positivamente a construção de uma postura antidiscriminatória face aos vieses introduzidos nas máquinas.

3.2. A construção social da ética como mecanismo de controle à discriminação algorítmica

Em sua famigerada obra "Justiça", o professor Michael J. Sandel apresenta ao leitor o dilema ético do bonde desgovernado:

> Suponha que você seja o motorista de um bonde desgovernado avançando sobre os trilhos a quase 100 quilômetros por hora. Adiante, você vê cinco operários em pé nos trilhos, com as ferramentas nas mãos. Você tenta parar, mas não consegue. Os freios não funcionam. Você se desespera porque sabe que, se atropelar esses cinco operários, todos eles morrerão. (Suponhamos que você tenha certeza disso.) De repente, você nota um desvio à direita. Há um operário também naqueles trilhos, mas apenas um. Você percebe que pode desviar o bonde, matando esse único trabalhador e poupando os outros cinco. [...] Agora considere outra versão da história do bonde. Desta vez, você não é o motorneiro, e sim um espectador, de pé numa ponte acima dos trilhos. (Desta vez, não há desvio.) O bonde avança pelos trilhos, onde estão cinco operários. Mais uma vez, os freios não funcionam. O bonde está prestes a atropelar os operários. Você se sente impotente para evitar o desastre - até que nota, perto de você, na ponte, um homem corpulento. Você poderia empurrá-lo sobre os trilhos, no caminho do bonde que se aproxima. Ele morreria, mas os cinco operários seriam poupados. (Você pensa em pular os trilhos, mas se dá conta de que é muito leve para parar o bonde.) Empurrar o homem pesado sobre os trilhos seria a coisa certa a fazer? (SANDEL, 2019, p. 30-31

Mais adiante, o autor alerta que os "dilemas morais têm origem em princípios morais conflitantes" (SANDEL, 2019, p. 33). Afinal, do ponto de vista ético e moral, qual a melhor ação a ser tomada: matar um inocente ou salvar um número maior de vidas?

Tal controvérsia torna-se ainda mais problemática ao se imaginar que, atualmente, uma máquina poderia ser introduzida em um cenário análogo ao idealizado pelo professor harvardiano, sendo ônus a ser suportado pelas produtoras de IA definir o rumo a ser seguido por suas respectivas máquinas, o que demandaria, inevitavelmente, um exame ético-filosófico sobre a hipótese.

Pensando nisso, o Massachusetts Institute of Technology (MIT), em parceria com outras instituições, produziu um *survey* global denominado "*The Moral Machine*", buscando opiniões da população sobre como achavam que a máquina deveria proceder em situações como a então apresentada, sendo identificado que, em que pese pautados em princípios morais universais, os entrevistados deram soluções diversas, a saber: de acordo com suas respectivas bases culturais. Para ilustrar, verificou-se que os entrevistados residentes em países asiáticos tiveram a inclinação de valorizar mais a vida de pessoas idosas em detrimento de jovens, enquanto os participantes ocidentes valorizaram mais a vida dos jovens (ALENCAR, 2021, p. 85).

Ainda nessa perspectiva, Ana Catarina Alencar discorre sobre a polêmica envolvendo um executivo da Mercedes-Benz, ao afirmar que os carros autônomos da empresa prefeririam a vida dos passageiros e condutor do automóvel às vidas de eventuais transeuntes:

> Em 2016, o executivo da Mercedes-Benz, Christoph von Hugo, foi envolvido em uma grande polêmica ao afirmar que os carros autônomos da montadora vão salvar o motorista e os passageiros do veículo, mesmo que isso signifique sacrificar a vida de pedestres em uma situação na qual essas sejam as únicas opções (ALENCAR, 2021, p.86)

Logo, o tema virou objeto de incansáveis debates no meio midiático, que, nas entrelinhas, se preocupou em analisar a eticidade das afirmações do executivo[6]. Sob esse prisma, identifica-se a imprensa como um importante instrumento de conscientização social e como propulsor da ética.

De mais a mais, verifica-se que princípios internacionais como os estabelecidos pela OCDE – expostos no início da presente pesquisa –, tendem a apresentar um conteúdo genérico, ao passo que visam a universalização de suas aplicações, o que acaba por desfavorecer a proteção dos direitos e interesses individuais e coletivos que poderiam ser afetados pelo uso de sistemas inteligentes (MARANHÃO; FLORÊNCIO; ALMADA, 2021, p. 162).

Para assegurar a aplicabilidade protetiva, faz-se necessário um exame ético e interdisciplinar dos princípios internacionalmente dispostos, a fim de conceber regras mais específicas para cada setor da sociedade (economia, medicina, direito etc), e, quiçá, auxiliar na posterior regulamentação setorizada da IA, tarefa que incumbe – dentre outros – à academia:

> Entendemos que o melhor caminho para discussões éticas voltadas à aplicabilidade seja a análise "*bottom-up*", buscando um equilíbrio reflexivo entre princípios gerais e casos concretos em setores específicos. Ou seja, trata-se de discutir não os princípios universais da ética computacional ou algorítmica, mas de desenhar princípios específicos para diferentes setores de aplicação: ética algorítmica no campo da medicina, no campo jurídico, no âmbito comercial etc. Refletindo as prioridades de cada domínio de aplicação, será possível construir sistemas inteligentes centrados nas demandas humanas [...]. O equilíbrio ideal dependerá não só das capacidades tecnológicas dos sistemas inteligentes, mas também das peculiaridades de cada domínio de

[6] Entre as matérias jornalísticas da época, destaca-se: RODRIGUES, Marcelo. Sim, o carro autônomo da Mercedes vai atropelar VOCÊ para salvar o condutor. **Tecmundo**. 15 out. 2016. Disponível em: https://www.tecmundo.com.br/mercedes/110591-sim-carro-autonomo-mercedes-atropelar-voce-salvar-condutor.htm. Acesso em 30 ago. 2025.

> aplicação e das exigências da lei. Nesse sentido, é salutar o desenvolvimento de uma literatura acadêmica brasileira sobre o tema que, em diálogo com o estado da arte da pesquisa global, busca enfrentar problemas no contexto brasileiro e nas condições do ordenamento pátrio, sem, contudo, perder de vista que muitos desses desafios também se fazem presentes em outros países (MARANHÃO; FLORÊNCIO; ALMADA, 2021, p. 162).

Portanto, tendo em vista a complexidade do estudo da ética algorítmica, é patente a necessidade de um estudo transdisciplinar da matéria, nos moldes a seguir expostos:

> O desastroso acirramento de disparidades e iniquidades pré-existentes [por exemplo, com a retroalimentação da seletividade penal] chamou atenção para o fato de que o exame de vieses algorítmicos requer uma abordagem transdisciplinar - que demandará a união de esforços por parte de juristas, sociólogos, filósofos, psicólogos, cientistas políticos, programadores, desenvolvedores, matemáticos e estatísticos -, pois apenas uma análise que tenha em vista fatores sociais, técnicos e jurídicos será capaz de capturar as peculiaridades dos vieses profundamente enraizados nos algoritmos que progressivamente moldam, delimitam, determinam ou influenciam as decisões nas sociedades modernas (WOJCIECHOWSKI; ROSA, 2021, p. 101).

Ainda nesse enfoque, frisa-se a importância da obediência ao princípio da transparência pela Administração Pública, uma vez que a publicidade dos algoritmos utilizados é necessária para fiscalizar a lisura da programação por parte dos diversos setores da sociedade.

Nesse plano, destaca-se:

> Pode-se dividir o processo de prestação de contas dos modelos em ao menos duas etapas, ainda que se queira preservar a integralidade de seus códigos propriamente ditos. Em um momento de pré-elaboração dos códigos, é fundamental que a sociedade em geral tenha ciência de quais aspectos da realidade serão levados em conta pelo modelo, o que é percebido nos dados que serão utilizados como inputs. Dessa forma, é possível, por exemplo, contestar os motivos que levaram a certos dados serem considerados pertinentes e as razões pelas quais outros não foram sequer considerados. Após escritos tais códigos, pode-se fiscalizá-los por meio da auditoria de seus resultados, momento em que são detectados eventuais vieses ou distorções. Assim, uma vez implementado um modelo e verificado que ele penaliza um determinado segmento social desproporcionalmente aos demais, sem justificativa, pode-se pleitear sua alteração ou seu desativamento. Para tanto, mais uma vez, é necessário que tais resultados estejam acessíveis ao público, especialmente àqueles que são diretamente afetados. Por fim, e talvez mais importante, é necessário que a própria ideia de concepção do algoritmo provenha de uma construção conjunta entre todas as partes interessadas, a saber, gestores, público em geral e desenvolvedores. A participação de diversos setores da sociedade na elaboração de algoritmos é essencial para que suas definições de sucesso e as tarefas em que são empregados sejam

> estabelecidas da forma mais benéfica ao maior número de pessoas ou, quando isso não for possível, implementados de forma a mitigar ao máximo seus malefícios (BOEING; ROSA, 2020, p. 89-90).

Por derradeiro, conforme enfatizado, a ética e moralidade se constroem socialmente de maneira natural, sendo que, ao contrário da lei – que surge muitas vezes como consequência de um paradigma ético contínuo – tais acepções surgem como "influxos na atividade avaliativa das pessoas, antes ou na eminência de agirem" (ZANON JÚNIOR, 2014, p. 16), não possuindo coercitividade.

Daí por que a regulamentação das novas tecnologias pelo Estado é urgente.

3.3. A utilização da ética no campo jurídico como instrumento de salvaguarda dos direitos humanos frente à inteligência artificial

Um ano após a recomendação da OCDE sobre o desenvolvimento tecnológico de algoritmos, o Conselho Nacional de Justiça (CNJ) elaborou a Resolução n. 332/2020, que, entre outras medidas, "dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário [...]" (BRASIL, 2020).

Depreende-se, pois, que o Poder Judiciário brasileiro não se olvida das inúmeras ocorrências de violações aos direitos humanos/fundamentais por intermédio de algoritmos discriminatórios.

Nessa planura, assenta-se no artigo 7º da referida norma:

> Art. 7º As decisões judiciais apoiadas em ferramentas de Inteligência Artificial devem preservar a igualdade, a não discriminação, a pluralidade e a solidariedade, auxiliando no julgamento justo, com criação de condições que visem eliminar ou minimizar a opressão, a marginalização do ser humano e os erros de julgamento decorrentes de preconceitos.
> § 1º Antes de ser colocado em produção, o modelo de Inteligência Artificial deverá ser homologado, a fim de identificar se preconceitos ou generalizações influenciaram seu desenvolvimento, acarretando tendências discriminatórias em seu funcionamento.
> § 2º Verificado viés discriminatório de qualquer natureza ou incompatibilidade do modelo de Inteligência Artificial com os princípios previstos nesta Resolução, deverão ser adotadas medidas corretivas.
> § 3º A impossibilidade de eliminação do viés discriminatório do modelo de Inteligência Artificial implicará na descontinuidade de sua utilização, com o consequente registro de seu projeto e as razões que levaram a tal decisão (BRASIL, 2020, grifou-se).

No âmbito legislativo – sabendo que as decisões expedidas por inteligências artificiais podem conter erros/vieses –, a Lei Geral de Proteção de Dados dispõe sobre a

garantia do titular de dados de ter a revisão de decisões tomadas por máquinas autônomas:

> Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base no tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir seu perfil pessoal, profissional, de consumo e de crédito, ou os aspectos de sua personalidade.
> § 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados na decisão automatizada, observados os segredos comerciais e industriais.
> § 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 (BRASIL, 2019, s. p.).

Com base no princípio da inafastabilidade da jurisdição, tais normas, acaso violadas, poderiam ser submetidas à análise de magistrados, sob o crivo do contraditório e da ampla defesa.

Demasiado, conforme exposto no decorrer da pesquisa, existem casos concretos que – se submetidos a julgamento perante o Poder Judiciário e adotada uma lógica dworkiana – certamente seriam recepcionados como "*hard cases*" (casos difíceis), isto é, que demandam um maior juízo interpretativo da lide pelo magistrado.

Concomitantemente a esse cenário, tem-se a ausência de maior regulamentação a respeito do assunto e, como é notório, magistrados não podem deixar de apreciar lides concretas levadas ao Judiciário, mesmo na ausência de normas. Assim, a ética dos dados seria, em primeira análise, um importante instrumento a ser utilizado pelos julgadores na análise de lides envolvendo inteligência artificial.

Sem embargo, verifica-se que a possibilidade de uso da ética/moral pelos magistrados não é ideia homogênea na doutrina. Por um lado, temos concepções como a de Manuel Atienza, que trata da moralidade como figura indispensável a ser levada em conta pelos juízes, não podendo ser desvinculada da Lei:

> O raciocínio jurídico não é completamente independente do de caráter moral. Tem, cabe dizer, uma autonomia relativa, no sentido de que, ainda que o raciocínio jurídico esteja submetido ao raciocínio moral, isso não quer dizer, que seja, como às vezes se supõe, um caso especial dele. Não o é, porque as regras da argumentação moral (racional) não se aplicam sempre (ou não todas elas) à argumentação jurídica em seus diversos contextos. O que é certo é que as regras que regem os diversos âmbitos da argumentação jurídica têm de estar justificadas do ponto de vista moral, isto é, têm de poder ser derivadas de razões morais. Isso quer

> dizer, por exemplo, que não parece ter sentido exigir de um advogado (em uma peça de defesa ou de acusação) uma sinceridade plena (essa seria uma das regras da argumentação moral), porém, entretanto, a possibilidade de ocultar informação quando se desempenha esse papel tem que poder ser justificada mediante o discurso moral (ATIENZA, 2013, p. 83).

Ainda, o autor complementa que a decisão, pautada em noções morais, pelo magistrado em casos difíceis, deve estar na linha da opinião majoritária do corpo social, visando justamente assegurar a aplicação da ética universalmente estabelecida. Veja-se:

> O apelo à moral social (a noções, valores, características da moral social) está incorporado de modo explícito, ocasionalmente, em normas jurídicas. Nesses casos, o uso desse critério para avaliar as decisões judiciais não suscita dúvidas (ainda que, naturalmente, possa ser discutível se o juiz interpretou bem ou não o correspondente critério da moral social). O que poderia parecer mais questionável é se esse critério pode ser usado (se tem força justificativa) quando as normas do Direito positivo não o tenham previsto, ou seja, se nos casos difíceis, quando se trata de optar por uma ou outra interpretação a respeito de determinado conceito valorativo (por exemplo, como entender a liberdade ou os limites da liberdade), os juízes devem decidir de acordo com a opinião majoritária das pessoas, se devem seguir (isto é, utilizar como premissa de seu argumento) o padrão marcado pela moral social e não a opinião que ele, como individuo, considere preferível (ATIENZA, 2013, p. 137).

Assim, a moralidade (individual) ou a ética (coletiva) servem, para o autor, como um critério valorativo a ser utilizado pelo magistrado quando da ausência de normas jurídicas.

Nessa linha, ensina Robert Alexy que a atividade jurisdicional não pode se esgotar na subsunção da norma ao fato:

> Em quase todos os escritos de metodologia, destaca-se atualmente que a Ciência do Direito e a jurisprudência não podem prescindir de tais valorações. Nesse sentido, Larenz fala do "reconhecimento de que a aplicação da lei não se esgota na subsunção, mas exige, em grande medida, valorações do aplicador". Muller considera que "uma Ciência do Direito e uma jurisprudência sem decisões nem valorações... (não seria) nem prática nem real. Esser constata que "as valorações... (têm) uma importância central, de algum modo problemática, em todas as decisões. Kriele chega à conclusão de que não se pode "desconhecer o elemento valorativo, normativo-teleológico e político-jurídico contido em toda interpretação. Engisch reconhece que "as próprias leis em todos os ramos do Direito (são) atualmente construídas de modo que os juízes e a Administração Pública não ordenam e fundamentam suas decisões por meio da subsunção a rígidos conceitos jurídicos, cujo conteúdo se revela, com certeza, por meio da interpretação, mas têm de valorar autonomamente e decidir como colegisladores (ALEXY, 2011, p. 23-24).

Luigi Ferrajoli diverge, aduzindo que o modelo adotado por Alexy e Atienza – denominado pelo autor "principialist" – pode violar o princípio da separação dos poderes, não admitindo, portanto, a ponderação pelos magistrados. Muito mais do que no modelo principialista de Ferrajoli - que confia a solução das aporias e dos conflitos entre direitos à ponderação judicial, enfraquecendo a normatividade das Constituições e a fonte de legitimação da jurisdição -, o paradigma garantista do constitucionalismo rígido exige que o Poder Judiciário seja o mais limitado e vinculado possível pela lei e pela Constituição, conforme o princípio da separação de poderes e a natureza quanto mais legítima mais cognitiva, e não discricionária, da jurisdição.

Os juízes, com base neste paradigma, não ponderam normas, mas sim as circunstâncias fáticas que justificam ou não a sua aplicação. Não lhes é dado criar nem ignorar normas, o que implicaria invasão no campo da legislação, senão censurar a invalidade destas por violação à Constituição, anulando-as no âmbito da jurisdição constitucional ou suscitando exceções de inconstitucionalidade no âmbito da jurisdição ordinária, intervindo, em ambos os casos, não na esfera legítima, mas na esfera ilegítima da política.

A legitimidade da jurisdição funda-se, nesse quadro teórico, no caráter mais cognitivo possível da subsunção na aplicação da lei, dependente do grau de taxatividade e de determinação da linguagem legal, bem mais do que de sua formulação como regra, sendo a indeterminação normativa e a consequente discricionariedade judicial sempre fatores de deslegitimação da atividade do juiz. O cognitivo judicial (veritas non auctoritas facit iudicium), mesmo como ideal regulativo, é a outra face do convencionalismo legal, isto é, do princípio juspositivista da legalidade (auctoritas non veritas facit legem). A satisfação de semelhante ideal é uma questão de grau, dependente do nível de determinação da linguagem legal. Ao ultrapassar certo grau de indeterminação, convertendo-se em poder criativo ou de disposição, o Poder Judiciário perde qualquer legitimidade, seja quando as normas forem expressas em forma de princípios, seja quando forem expressas em forma de regras (FERRAJOLI, 2022, p. 53-54).

Dessa forma, detecta-se uma divergência doutrinária quanto a (im)possibilidade de utilização da ética dos dados como mecanismo de ponderação por parte dos magistrados quando da análise de lides envolvendo inteligência artificial, o que demanda maior aprofundamento da presente pesquisa.

Considerações Finais

Analisando a versada conjuntura de fatores, infere-se que a concepção de ética jamais se dará por inerte, tampouco sobreviverá às barreiras do tempo e às mutações sociais. Em verdade, continuamente se transformará no decorrer da evolução do ser humano, sendo inconcebível se pensar em uma única definição precisa a respeito desse termo filosófico, que tende a ser desenvolvido espontaneamente, com o debate popular a respeito de controvérsias que chocam e indignam a população – como no exposto caso da Beauty IA –, ou diante do anseio da comunidade, quando ultrajada, em ter a garantia da sanção aos responsáveis por ações moralmente questionáveis – como no abalizado caso do bonde desgovernado pensado sobre a perspectiva da automação veicular.

Constata-se, pois, que a sociedade já está se posicionando de forma crítica às mudanças ocasionadas pelo uso da inteligência artificial, podendo se falar da mídia e do meio acadêmico como importantes saídas sociológicas para obstar a expansão desenfreada de mecanismos que podem comprometer direitos humanos e fundamentais da população que, cada vez mais, se valerá do uso da inteligência artificial em seu cotidiano. Nesse sentido, a análise transdisciplinar do tema também se faz necessária, sobretudo considerando a futura possibilidade de maior regulamentação da temática.

Não obstante, sendo as questões envolvendo inteligência artificial levadas à apreciação jurisdicional, constatou-se divergência doutrinária no tocante à (im)possibilidade do uso da ética dos dados como instrumento jurídico apto a embasar julgamentos realizados por magistrados, porquanto tal postura pode ferir a separação dos poderes e/ou enfraquecer a eficácia do texto constitucional. Assim, neste ponto, a pesquisa restou inconclusiva e demanda, portanto, estudos adicionais sobre o tema.

Referências

ALENCAR, Ana Catarina. **Inteligência Artificial, Ética e Direito**. Saraiva Educação SA, 2022. Disponível em: https://bibliotecadigital.saraivaeducacao.com.br/epub/798038?title=Inteligência%20Artificial%20e%20Direito. Acesso em 30 mar. 2026.

ALEXY, Robert. **Teoria da argumentação jurídica**: a teoria do discurso racional como teoria da fundamentação jurídica. 3. ed. Rio de Janeiro: Forense, 2011.

ANGELINI, Roberta. **Intelligenza Artificiale e governance. Alcune riflessioni di sistema.** *In*: PIZZETTI, Franco (org.). **Intelligenza Artificiale, Protezione dei dati personali e regolazione.** Torino: Giappichelli, 2018.

ATIENZA, Manuel. **Curso de argumentação jurídica**. 1. ed. Curitiba: Alteridade, 2017.

BASIGLI, Nicolò. **La possibilità della repubblicanizzazione dell'intelligenza artificiale:** la necessità del diritto transnazionale per la protezione dei diritti fondamentali della persona umana. Tese de doutorado, Università degli Studi di Perugia / UNIVALI, 2022.

BOEING, Daniel Henrique Arruda; ROSA, Alexandre Morais da. **Ensinando um robô a julgar**: pragmática, discricionariedade, heurísticas e vieses no uso de aprendizado de máquina no judiciário. 1ª ed. Florianópolis: Emais Academia, 2020.

BOSTROM, Nick. **Superinteligência**: caminhos, perigos, estratégias. Rio de Janeiro: Intrínseca, 2015.

BRASIL. LEI Nº 13.709, DE 14 DE AGOSTO DE 2018. **Lei Geral de Proteção de Dados Pessoais (LGPD)**. Brasília, DF. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em 30 mar. 2026.

BRASIL. Resolução Nº 332 de 21/08/2020. **Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências**. Disponível em: https://atos.cnj.jus.br/atos/detalhar/3429. Acesso em 30 mar. 2026.

CORTELI, Anna Carolina Nunes et al. **Brasil a caminho da OCDE**: explorando novos desafios. Coordenadores: Vera Thorstensen; Thiago Rodrigues São Marcos Nogueira – São Paulo: VT Assessoria Consultoria e Treinamento Ltda. 2020.

EUROPA-OCDE. **Recomendation of the Council on Artificial Intelligence**. OECD/LEGAL/0449. Disponível em: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjlssTo1_H7AhW1jZUCHUFgAXQQFnoECAwQAQ&url=https%3A%2F%2Flegalinstruments.oecd.org%2Fapi%2Fprint%3Fids%3D648%26lang%3Den&usg=AOvVaw3bU62HpvCxeAcd6gxRGeJ6. Acesso em 30 mar. 2026.

FERRAJOLI, Luigi et al. **Garantismo, hermenêutica e (neo)constitucionalismo**: um debate com Luigi Ferrajoli. 2. ed. ampl. São Paulo: Tirant lo Blanch, 2022.

FORNASIER, Matheus Oliveira. **Inteligência artificial e democracia**: oportunidades e desafios. NOMOS: Revista do Programa de Pós-Graduação em Direito da UFC, Fortaleza, v.41, n.1, jan./jun.2021, p.171-194.

GARCIA, Ana Cristina. **Ética e Inteligência artificial**. Revista Computação Brasil, Rio de Janeiro, n. 43, nov. 2020, p. 1-22. Disponível em: https://sol.sbc.org.br/journals/index.php/comp-br/article/view/1791. Acesso em 30 ago. 2025.

GOVERNMENT OFFICE FOR SCIENCE. Artificial Intelligence: opportunities and implications for the future of decision making. London: **GO-Science**, 2015. Disponível em: https://www.gov.uk/go-science . Acesso em 30 mar. 2026.

MACHADO SEGUNDO, Hugo de Brito. **Direito e Inteligência artificial** [recurso eletrônico]: o que os algoritmos têm a ensinar sobre interpretação, valores e justiça. Indaiatuba, SP: Editora Foco, 2023. Epub. Disponível em: https://books.google.com.br/books?id=OFyYEAAAQBAJ&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. Acesso em 30 mar. 2026.

MARANHÃO, Juliano Souza de Albuquerque; FLORÊNCIO, Juliana Abrusio; ALMADA, Marco. **Inteligência artificial aplicada ao direito e o direito da inteligência artificial**. Suprema: revista de estudos constitucionais, Brasília, v. 1, n. 1, p. 154-180, jan./jun. 2021.

MARMELSTEIN, George. **Testemunhando a injustiça**: a ciência da prova testemunhal e das injustiças inconscientes. São Paulo: Editora Juspodivm, 2022.

MITCHELL, Tom. **Machine Learning**. New York: McGraw-Hill Science/Engineering/Math, 1997.

OECD. **OECD AI Policy Observatory**, 2019. G20 AI Principles. Disponível em: https://oecd.ai/en/wonk/documents/g20-ai-principles. Acesso em 30 mar. 2026.

PIMENTEL, A. F.; SOUTO ORENGO, B. **Perspectivas de aplicação da inteligência artificial no direito processual**: análise sobre as diretrizes éticas e eficiência jurisdicional. Revista Brasileira de Sociologia do Direito, v. 8, n. 3, p. 305-325, 31 ago. 2021.

PIZZETTI, Franco. **Intelligenza Artificiale, Protezione dei dati personali e regolazione.** Torino: Giappichelli Editore, 2018.

SAMUEL, Arthur L. Some Studies in Machine Learning Using the Game of Checkers. **IBM Journal of Research and Development,** v. 3, n. 3, 1959.

SANDEL, Michael J. **Justiça**: O que é fazer a coisa certa? Tradução de Heloísa Matias e Maria Alice Máximo. 27ª ed. Rio de Janeiro: Civilização Brasileira, 2019.

SEARLE, John. Minds, Brains and Programs. Behavioral and Brain Sciences. **Cambridge University Press**. v. 3, n. 3, 1980. Disponível em: https://www.cambridge.org/core/journals/behavioral-and-brain-sciences/article/abs/minds-brains-and-programs/DC644B47A4299C637C89772FACC2706A . Acesso em 30 mar. 2026..

SOARES, R. M. F. **Teoria Geral do Direito**. 5. ed. São Paulo: Saraiva, 2019. E-book.

TAKAKURA, Flávio Iassuo; DUARTE, Luciana Gaspar Melquíades. **Inteligência Artificial no Direito**: Dilemas e Contribuições. Revista de Direito, Inovação, Propriedade Intelectual e Concorrência, v. 8, n. 1, p, 1-23, 11 set. 2022.

UNESCO. **Recomendação sobre a Ética da Inteligência Artificial**. Paris: UNESCO, 2021. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137. Acesso em 30 mar. 2026.

UNIÃO EUROPEIA. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, relativo à Inteligência Artificial e que altera determinados atos legislativos da União (AI Act).** Bruxelas, Parlamento Europeu e Conselho da União Europeia, 13 jun. 2024. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689 . Acesso em 30 mar. 2026.

WOJCIECHOWSKI, Paola Bianchi; ROSA, Alexandre Morais da. **Vieses da Justiça:** como as heurísticas e vieses operam nas decisões penais e a atuação contraintuitiva. Florianópolis: Emais Editora, 2021.

ZANON JUNIOR, Orlando Luiz. **Moral, Ética e Direito**. Revista da Escola Superior da Magistratura do Estado de Santa Catarina (ESMESC), v. 21, n. 27, 2014.