

UMA ANÁLISE DOS ALGORITMOS E DOS VIESES ESCONDIDOS NA IA: OS DESAFIOS ÉTICOS E JURÍDICOS INERENTES

AN ANALYSIS OF ALGORITHMS AND HIDDEN BIAS IN ARTIFICIAL INTELLIGENCE: THE INHERENT ETHICAL AND LEGAL CHALLENGES

ANTONIO FELIPE DE SOUSA¹
RENATA ALBUQUERQUE LIMA²

RESUMO

A ascensão da Inteligência Artificial (IA) tem gerado importantes discussões sobre os desafios éticos e jurídicos, particularmente devido aos vieses algorítmicos que podem influenciar decisões autônomas. Esta pesquisa, de natureza documental bibliográfica, explora a complexidade desses vieses, apresentando casos emblemáticos como o algoritmo COMPAS, que mostrou viés racial em previsões de reincidência criminal, e o Beauty.AI, que perpetuou padrões de beleza discriminatórios, e o Chatbot Tay que, em poucas horas de uso, apresentou conteúdo indesejado. A falta de transparência nos algoritmos (deep learning) é um problema central, dificultando a contestação de decisões injustas e levantando questões sobre responsabilidade. Para mitigar esses problemas, propõe-se a necessidade de uma maior transparência e explicabilidade dos algoritmos, além de uma regulamentação rigorosa que assegure justiça e equidade. A colaboração entre diversas áreas do conhecimento é essencial para desenvolver soluções que promovam uma sociedade mais justa, onde a IA possa ser utilizada de forma ética e responsável.

Palavras-chave: vieses algorítmicos; falta de transparência; *deep learning*; ética e responsável.

ABSTRACT

The rise of Artificial Intelligence (AI) has generated important discussions about ethical and legal challenges, particularly due to algorithmic biases that can influence autonomous decisions. This research, of a bibliographical documentary nature, explores the complexity of these biases, presenting emblematic cases such as the COMPAS algorithm, which showed racial bias in predictions of criminal recidivism, and Beauty.AI, which perpetuated discriminatory beauty standards, and the Chatbot Tay, which, within a few hours of use, presented unwanted

1 Mestrando em Filosofia (UVA), Professor Especialista em Neuropsicopedagogia (FASOUZA), Pedagogo (UVA), Acadêmico do 6º semestre do Curso de Direito da Faculdade Luciano Feijão (FLF). Presidente do Instituto Renascer de Luz. E-mail: professorantoniofelipe@hotmail.com. Lattes: <https://lattes.cnpq.br/2104305957238308>. Orcid: <https://orcid.org/0000-0002-7820-6397>.

2 Pós-Doutora em Direito pela Universidade Federal de Santa Catarina (UFSC). Doutora em Direito Constitucional pela Universidade de Fortaleza (UNIFOR). Professora Adjunta do Curso de Direito da Universidade Estadual Vale do Acaraú – UVA. Professora da Graduação e do Mestrado em Direito da UNICHRISTUS. Coordenadora do Curso de Direito da Faculdade Luciano Feijão. E-mail: realbuquerque@yahoo.com. Orcid: <https://orcid.org/0000-0002-4019-9558>.

Como citar esse artigo:/How to cite this article:

SOUZA, Antonio Felipe de; LIMA, Renata Albuquerque. Uma análise dos algoritmos e dos vieses escondidos na IA: os desafios éticos e jurídicos inerentes. **Revista Meritum**, Belo Horizonte, v. 19, n. 2, p. 60-74, 2024. DOI: <https://doi.org/10.46560/meritum.v19i2.10149>.

content. The lack of transparency in algorithms (deep learning) is a central problem, making it difficult to challenge unfair decisions and raising questions about accountability. To mitigate these problems, we propose the need for greater transparency and explainability of algorithms, in addition to strict regulation that ensures justice and equity. Collaboration between different areas of knowledge is essential to develop solutions that promote a more just society, where AI can be used ethically and responsibly.

Keywords: algorithmic biases; lack of transparency; deep learning; ethical and responsible.

1. INTRODUÇÃO

A questão da Inteligência Artificial (IA) desponta como um dos tópicos mais candentes na atualidade. Ainda que possa parecer uma inovação dos tempos modernos, a ideia subjacente remonta ao século XX. Em 1936, o matemático britânico Alan Turing³ concebeu o que viria a ser conhecido como a “máquina de Turing” – um dispositivo teórico que lançou os fundamentos para o que hoje é a representação algorítmica. . Turing, com sua visão previdente, cita: “Por ora, direi apenas que a justificação reside no fato de que a memória humana é necessariamente limitada. ” (Turing, 1936, p. 231). A proposta de Turing abriu portas para um campo de estudo que, décadas depois, tomaria o nome de Inteligência Artificial.

Foi em 1956 que o professor John McCarthy, da Universidade de Dartmouth nos EUA, cunhou o termo “inteligência artificial” e reuniu uma comunidade de cientistas para explorar o conceito de máquinas capazes de aprendizado. A partir desse ponto, uma série de marcos históricos delinearam a trajetória da IA. Em 1966, viu-se o surgimento do primeiro chatbot, um embrião das interações homem-máquina que se tornariam corriqueiras. Avançando para 1972, a área médica adotou a IA para diagnósticos complexos, assinalando um passo importante na aplicação prática dessa tecnologia.

Conforme os registros disponibilizados pela plataforma CODEX⁴, o Brasil alcançou um feito de considerável relevância jurídica em novembro de 2022: a nação ultrapassou a significativa marca de 100 milhões de processos judiciais armazenados, delineando um panorama impactante e colossal do cenário jurídico do país. O Brasil emerge como um país com uma notável cultura de litigância, onde o ato de ajuizamento de ações tornou-se uma parte intrínseca da dinâmica social e legal. Esse cenário reflete não apenas a busca por resoluções justas, mas também as complexidades de um ambiente em constante transformação, onde os cidadãos buscam amparo legal para uma ampla gama de assuntos, desde questões pessoais até contendas mais abrangentes.

Destaquemo-nos como um ponto positivo na utilização da inteligência artificial para tarefas rotineiras e repetitivas no contexto jurídico, como a leitura de petições repetitivas, tal feito já se realiza no Brasil e foi promovido pelo STF (Supremo Tribunal Federal) através do denominado “Projeto Victor”, em homenagem ao ilustre Ministro Victor Nunes Leal⁵ de desde 2018 através

3 Biografia de Alan Turing. Disponível em: https://www.ebiografia.com/alan_turing/. Acesso em 20 ago. 2023.

4 CODEX atinge mais de 100 milhões de processos judiciais armazenados. Disponível em: <https://www.cnj.jus.br/codex-atinge-mais-de-100-milhoes-de-processos-judiciais-armazenados/>. Acesso em 20 ago. 2023.

5 Disponível em: <https://www.oab.org.br/noticia/60153/juristas-que-marcaram-a-historia-do-pais-victor-nunes-leal#:~:text=Victor%20Nunes%20Leal%20foi%20procurador,Rep%C3%A3o%20do%20governo%20Juscelino%20Kubitschek>. Acesso em: 20 abr. 2024.

da Ministra Cármem Lúcia⁶. No entanto, ao considerarmos essa abordagem, é importante refletir sobre um possível aspecto negativo: a potencial perda da dimensão humana na análise das complexidades legais e na interpretação das situações individuais. Afinal, por mais que essa abordagem pudesse acelerar o andamento dos processos e impulsionar a agilidade da justiça, há um sentimento compartilhado por muitos de que a empatia, a compreensão contextual e o discernimento humano desempenham um papel vital no sistema jurídico.

Além disso, surge uma preocupação amplamente discutida: a inserção inadvertida de vieses preconceituosos e desigualdades presentes nos dados utilizados para treinar os algoritmos de inteligência artificial. Como sabemos, as máquinas aprendem a partir de informações passadas e podem perpetuar, de forma involuntária, essas tendências discriminatórias. Nesse sentido, o uso da inteligência artificial no âmbito jurídico levanta questionamentos essenciais sobre justiça social, igualdade e imparcialidade.

Apresentaremos e discutiremos os casos de vieses algorítmicos ilustram claramente como esses problemas éticos se manifestam na prática. Em 2016, o concurso de beleza Beauty.AI⁷ utilizou um sistema de IA para avaliar os participantes com base em critérios objetivos, como simetria facial e uniformidade da pele. No entanto, o resultado final mostrou um viés claro: quase todos os vencedores eram brancos. Esse caso ilustra como os dados de treinamento, baseados predominantemente em imagens de atores brancos de Hollywood, influenciaram negativamente o algoritmo, perpetuando um padrão de beleza que exclui minorias.

Outro exemplo é o sistema COMPAS⁸ (Correctional Offender Management Profiling for Alternative Sanctions), utilizado nos tribunais dos EUA para avaliar o risco de reincidência de criminosos. Estudos mostraram que o COMPAS tende a superestimar o risco de reincidência de criminosos negros e subestimar o de criminosos brancos, evidenciando um viés racial significativo. Esse caso destaca a gravidade dos vieses algorítmicos em decisões judiciais, com profundas implicações éticas e sociais.

O chatbot Tay⁹ da Microsoft, lançado em 2016, foi projetado para interagir com usuários no Twitter e aprender com essas interações. Em poucas horas, Tay começou a gerar conteúdo ofensivo e racista, refletindo o comportamento dos usuários com quem interagia. Esse incidente sublinha os perigos de expor algoritmos de IA a dados não moderados e a necessidade de mecanismos robustos para evitar a propagação de conteúdo prejudicial.

A análise dos algoritmos e dos vieses escondidos na IA revela um problema ético e moral significativo que exige atenção urgente. É preciso abordar esses desafios, enfatizando a necessidade de justiça, transparência e responsabilidade. Os casos discutidos ilustram a complexidade dos vieses algorítmicos e suas consequências, reforçando a importância de um escrutínio rigoroso e de práticas éticas no desenvolvimento e implementação da IA.

É fundamental implementar mecanismos de auditoria e revisão contínua dos sistemas de IA, assegurando que eles evoluam para eliminar vieses e promover a justiça. A educação e a

6 Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=388443&ori=1>. Acesso em: 20 abr. 2024.

7 Disponível em: <https://www.theguardian.com/technology/2016/sep/08/artificial-intelligence-beauty-contest-doesnt-like-black-people>. Acesso em: 16 abr. 2024.

8 Disponível em: <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 18 abr. 2024.

9 Disponível em: https://brasil.elpais.com/brasil/2016/03/24/tecnologia/1458855274_096966.html. Acesso em: 18 abr. 2024.

sensibilização sobre os vieses algorítmicos também são essenciais para que desenvolvedores e usuários compreendam as implicações éticas de suas decisões.

A transformação digital deve ser orientada por princípios éticos que garantam a inclusão e a equidade. A colaboração entre diferentes áreas do conhecimento é crucial para desenvolver soluções que promovam uma sociedade mais justa e igualitária, onde a tecnologia seja um instrumento de progresso para todos.

2. METODOLOGIA

Esta pesquisa utilizou o método documental bibliográfico, que é caracterizado pela análise de documentos e fontes teóricas já publicadas, com o intuito de compreender e aprofundar o tema investigado. Conforme assevera Marconi e Lakatos (2017): “*Dessa forma, a pesquisa bibliográfica não é mera repetição do que já foi dito ou escrito sobre certo assunto, visto que propicia o exame de um tema sob novo enfoque ou abordagem, chegando a conclusões inovadoras.*” (Marconi; Lakatos, 2017, p. 200). Essa abordagem metodológica permite não apenas uma síntese de informações existentes, mas também uma proposição de novos ângulos interpretativos, possibilitando um exame crítico e inovador.

Complementando essa visão, Gil (2002) destaca que a pesquisa bibliográfica é essencial para sistematizar e categorizar conhecimentos existentes, permitindo ao pesquisador identificar lacunas e avanços no campo de estudo. Para Gil, essa abordagem possibilita o entendimento aprofundado de fenômenos a partir de uma base sólida de informações já publicadas. Nesse sentido, a análise de fontes secundárias permite não apenas a revisão do estado da arte sobre o tema, mas também o desenvolvimento de novas hipóteses e interpretações.

No contexto desta pesquisa, o método bibliográfico foi instrumental para analisar os desafios éticos e jurídicos relacionados à inteligência artificial (IA), com abordagem específica em visões algorítmicas. A partir dessa abordagem, foram investigados casos práticos emblemáticos, que ilustram as complexidades e os impactos éticos de sistemas automatizados quando submetidos a dados enviados. O método bibliográfico foi de fundamental importância para analisar os desafios éticos e jurídicos relacionados à inteligência artificial (IA). Esses estudos de caso, amplamente documentados em literatura acadêmica e relatórios técnicos, foram essenciais para compreender as implicações sociais, éticas e jurídicas do uso de IA em decisões automatizadas.

Uma pesquisa bibliográfica bem fundamentada, atenua um papel crucial na integração de teorias e práticas, permitindo uma análise crítica das inter-relações entre características sociais e tecnológicas. No caso dos vieses algorítmicos, uma análise documental revela como os padrões históricos de discriminação são reproduzidos e amplificados por sistemas de IA. Exemplos ilustram como uma pesquisa documental pode contribuir para uma análise crítica e detalhada das implicações de sistemas de IA, destacando a necessidade de regulamentações éticas e jurídicas. O uso de métodos bibliográficos nos permite a construção de um panorama abrangente sobre o tema, promovendo uma visão integradora que abriga múltiplas dimensões do problema investigado.

No que diz respeito aos desafios metodológicos, Bardin (2016) enfatiza a importância de uma análise rigorosa e sistemática das fontes documentais, considerando a qualidade, a relevância e a diversidade dos materiais utilizados. Segundo essa orientação, esta pesquisa se baseia em literatura acadêmica, relatórios institucionais e estudos de caso, garantindo um exame criterioso e multifacetado das visões algorítmicas.

Além disso, a pesquisa documental bibliográfica não se limita à descrição dos problemas, mas avança na proposição de soluções. No caso dos vieses algorítmicos, as recomendações incluem a implementação de mecanismos de transparência e explicabilidade nos sistemas de IA, além de uma colaboração interdisciplinar que reúne especialistas em tecnologia, direito, filosofia e ciências sociais.

Concluindo, o método documental bibliográfico, como destacado por diversos autores, oferece um arcabouço robusto para análises críticas e inovadoras. A aplicação desse método nesta pesquisa foi essencial para desenvolver as interseções entre tecnologia, ética e justiça, propondo caminhos para um uso mais responsável e inclusivo da inteligência artificial.

3. CONSIDERAÇÕES SOBRE A INTELIGÊNCIA ARTIFICIAL

Para aprofundar a análise sobre os desafios éticos e morais relacionados à Inteligência Artificial (IA), é crucial considerar o impacto das decisões automatizadas nas diversas esferas sociais e econômicas. A utilização de IA em áreas como saúde, educação, segurança pública e mercado de trabalho traz à tona preocupações que vão além dos vieses algorítmicos, abrangendo questões de responsabilidade, transparência e direitos humanos.

A questão da responsabilidade é central na discussão sobre IA. Quando sistemas de IA tomam decisões que afetam a vida das pessoas, surge a pergunta: quem é responsável por essas decisões? A falta de clareza sobre a responsabilidade pode levar a uma “diluição da responsabilidade”, onde desenvolvedores, usuários e fabricantes se esquivam de suas responsabilidades, resultando em injustiças e danos não reparados. É essencial desenvolver um arcabouço legal e ético que clarifique a responsabilidade em casos de falhas ou danos causados por sistemas de IA. Este arcabouço deve incluir diretrizes claras para o desenvolvimento e o uso de IA, bem como mecanismos de reparação para vítimas de decisões injustas ou prejudiciais.

A inserção cada vez mais profunda da Inteligência Artificial na sociedade traz consigo um espectro de desafios éticos e questões legais intrincadas. À medida que a IA se imiscui em nossas vidas, é inevitável questionar o impacto que ela acarreta. Entre as inquietações prementes, despontam preocupações sobre o emprego, à medida que a automação ameaça determinadas ocupações, levando à necessidade de políticas de requalificação. É válido salientar que nesse contexto os mais vulneráveis estão sendo atingidos e sua regulação necessita ser realizada com urgência, conforme defende o Conselho de Comunicação Social (CSS) do Congresso Nacional e o Projeto de Lei (PL) 2.338/2023¹⁰.

10 Regulação de inteligência artificial é defendida no Conselho de Comunicação. Disponível em: <https://www12.senado.leg.br/noticias/materias/2023/06/05/regulacao-de-inteligencia-artificial-e-defendida-no-conselho-de-comunicacao>. Acesso em: 20 ago. 2023. Este projeto de Lei, representa um passo significativo nesse contexto, propondo um marco regulatório específico para a inteligência artificial.

A ausência de transparência dos algoritmos em sistemas de decisão é um problema significativo, especialmente no contexto jurídico. Os algoritmos são frequentemente considerados caixas-pretas, onde as entradas são conhecidas, mas o processo de cálculo que leva às saídas (resultados) não é claro. Isso levanta questões sérias sobre a justiça e a equidade dos sistemas automatizados.

No trecho abaixo, citado de Nunes e Marques (2018), os autores destacam a dificuldade de defesa contra um “índice” cujo método de cálculo é desconhecido. Sem entender como as respostas ou dados fornecidos são processados para gerar o resultado final, os acusados não conseguem contestar ou verificar a precisão e a justiça desse resultado. Esse cenário impede a aplicação do devido processo constitucional, que exige transparência e a possibilidade de contestação.

A ausência de transparência do algoritmo é especialmente crítica nesse caso. Como defender-se de um “índice” sem saber o método de seu cálculo? Como submeter o “índice” ao controle do devido processo constitucional? Por mais que sejam divulgadas as perguntas realizadas, os acusados não sabem como suas respostas influenciam no resultado final (output). Dessa forma, a defesa do acusado torna-se impossibilitada por dados matemáticos opacos e algorítmicamente enviesados, mas camuflados, pela “segurança” da matemática, como supostamente imparciais, impessoais e justos. (Nunes; Marques, 2018, p. 7).

A opacidade dos algoritmos coloca os acusados em uma posição desvantajosa, pois eles não conseguem avaliar ou contestar os métodos utilizados para gerar os índices que os afetam diretamente. Isso levanta sérias preocupações éticas e legais sobre o uso de algoritmos em contextos que demandam alta responsabilidade e justiça, como o sistema jurídico.

Um exemplo preocupante é “indivíduos com baixo risco de reincidência¹¹” (Nunes; Marques, 2018, p. 6). Mesmo esses indivíduos podem ser prejudicados por decisões baseadas em algoritmos opacos, uma vez que não têm meios de contestar ou entender plenamente a avaliação que determina seu risco. Isso demonstra como a falta de transparência pode afetar a vida das pessoas de maneiras significativas e potencialmente injustas.

A solução para essa questão envolve a exigência de maior transparência e explicabilidade dos algoritmos. Os sistemas automatizados devem ser auditáveis e compreensíveis, permitindo que os indivíduos afetados possam entender e contestar as decisões tomadas. Além disso, é fundamental que haja uma regulamentação que assegure a justiça e a equidade nos processos algorítmicos, prevenindo e corrigindo vieses que possam surgir.

Os algoritmos que determinam decisões importantes, devem ser transparentes e compreensíveis para aqueles que são afetados por essas decisões. A opacidade dos algoritmos, muitas vezes chamada de “caixa preta algorítmica”, impede que as pessoas compreendam como e por que determinadas decisões foram tomadas, dificultando a contestação de decisões injustas. Esses enviesamentos precisam ser combatidos. A transparência algorítmica deve ser promovida através de regulamentos que exijam explicações claras e acessíveis sobre como os algoritmos funcionam e quais dados são utilizados.

Segundo Cortiz (2021) a preocupação com a transparência e a compreensão dos algoritmos de inteligência artificial é acentuada pelo fato de que muitos deles, especialmente aqueles que utilizam aprendizado profundo “deep learning”, operam de maneira tão complexa que se

11 Grifos nossos.

tornam praticamente ininteligíveis para a maioria dos usuários. No aprendizado de máquina tradicional “*machine learning*”, há uma estrutura estatística relativamente simples que conecta a entrada de dados “*input*” com a saída “*output*”. Isso significa que, embora o processo de aprendizado seja complexo, é possível rastrear e entender como as decisões são tomadas com base nos dados fornecidos.

No entanto, o “*deep learning*” adiciona uma camada extra de complexidade. Este tipo de aprendizado utiliza múltiplas camadas de redes neurais artificiais, que se sobrepõem e interagem de maneiras complicadas. Essas camadas são projetadas para imitar o funcionamento do cérebro humano, permitindo que o algoritmo aprenda a reconhecer padrões e tomar decisões com base em grandes quantidades de dados. Cada camada da rede processa os dados de uma maneira diferente e passa as informações para a próxima camada, criando um processo de tomada de decisão que é extremamente difícil de seguir e entender.

Essa complexidade torna-se um problema significativo porque, apesar dos avanços na área, ainda existem poucas ferramentas e técnicas eficazes para explicar como essas redes neurais chegam às suas conclusões. Isso significa que, mesmo especialistas em inteligência artificial podem ter dificuldades para decifrar as decisões de um algoritmo de “*deep learning*”. A falta de transparência pode levar a desconfiança e incertezas, especialmente quando essas tecnologias são aplicadas em áreas sensíveis, como diagnósticos médicos, decisões judiciais ou sistemas de segurança.

O trabalho de Cortiz (2021) destaca essa limitação, apontando para a necessidade urgente de desenvolver métodos melhores para entender e interpretar os processos internos dos algoritmos de “*deep learning*”. Sem essas ferramentas, é difícil garantir que esses sistemas estejam funcionando de maneira justa, ética e sem vieses ocultos.

3.1 UMA ANÁLISE DE PROGRAMAS DE INTELIGÊNCIAS ARTIFICIAIS



Fonte: <https://play.google.com/store/apps/details?id=com.app.tayapp&hl=pt>.
Acesso em: 22 jun. 2024.

O chatbot Tay, desenvolvido pela Microsoft, foi uma tentativa ambiciosa de capturar o interesse dos millennials através de interações no Twitter. Lançado em março de 2016, Tay foi projetado para aprender e evoluir a partir das conversas com usuários, adotando uma abordagem “casual e brincalhona” para se conectar com jovens de 18 a 24 anos. No entanto, a falta de robustez no design do sistema tornou Tay vulnerável a manipulações maliciosas. Em poucas horas, usuários começaram a bombardear Tay com conteúdo racista, sexista e xenófobo, resultando em respostas que endossavam teorias da conspiração, negavam o Holocausto e faziam comentários ofensivos, como chamar uma mulher de “puta estúpida” e apoiar a construção de um muro na fronteira com o México, ecoando a retórica de Donald Trump na época.

Esse incidente com Tay evidencia um problema crítico no desenvolvimento de IA: a incapacidade de prever e mitigar interações adversas e a propagação de conteúdo nocivo. A Microsoft teve que rapidamente desativar Tay, alegando que o sistema estava sendo desligado para “absorver” as lições de seu “dia ocupado”. Esta falha destaca a necessidade de mecanismos de controle mais robustos e de moderadores humanos que possam intervir quando os sistemas de IA são usados de maneira prejudicial. A resposta da Microsoft, que incluiu a remoção dos tweets ofensivos de Tay e uma promessa de ajustes futuros, sublinha a importância de um design cuidadoso que leva em conta as possíveis consequências negativas das interações em tempo real na internet.



Fonte: <https://juristas.com.br/revistajuristas/inteligencia-artificial/>. Acesso em: 22 jun. 2024.

Com base no artigo “Como analisamos o algoritmo de reincidência COMPAS¹²” de Jeff Larson, Surya Mattu, Lauren Kirchner e Julia Angwin apresenta uma investigação crítica sobre o algoritmo COMPAS, uma ferramenta amplamente utilizada para prever a reincidência criminal. O estudo se concentrou em examinar a precisão e possíveis vieses raciais do algoritmo, especialmente entre réus negros e brancos. Utilizando um conjunto de dados de mais de 10.000

12 Disponível em <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>. Acesso em: 22 jun. 2024.

réus no Condado de Broward, Flórida, a análise revelou que o COMPAS tende a classificar incorretamente réus negros como de alto risco de reincidência mais frequentemente do que réus brancos, enquanto os réus brancos são mais frequentemente subestimados em seu risco.

Os dados obtidos pela ProPublica¹³ mostraram que a ferramenta COMPAS previu corretamente a reincidência geral em 61% dos casos, mas teve uma precisão significativamente menor (20%) ao prever reincidência violenta. Além disso, o estudo destacou disparidades significativas: réus negros que não reincidiram tinham quase o dobro de probabilidade de serem classificados erroneamente como de alto risco em comparação com réus brancos; enquanto réus brancos que reincidiram foram erroneamente rotulados como de baixo risco quase duas vezes mais frequentemente do que reincidentes negros. Mesmo controlando por fatores como histórico criminal, idade e gênero, réus negros apresentavam uma probabilidade 45% maior de receberem pontuações de risco mais altas.

O artigo também discute estudos anteriores e o contexto mais amplo das ferramentas de avaliação de risco nos Estados Unidos. A análise crítica incluiu comparações com outros estudos que examinaram diferentes ferramentas de avaliação de risco, mostrando que muitas vezes falta um exame rigoroso da validade preditiva e dos possíveis vieses raciais. O estudo realizado pela ProPublica sugere que, embora o COMPAS tenha algum valor preditivo, ele também apresenta falhas significativas que podem exacerbar desigualdades raciais no sistema de justiça criminal. A análise sublinha a necessidade de revisões independentes e regulares dessas ferramentas para garantir que não perpetuem injustiças sistêmicas.



Fonte: <https://beauty.ai/>. Acesso em: 22 jun. 2024.

De acordo Ding (2023) discute o uso de ferramentas de inteligência artificial na moda e como essas tecnologias podem perpetuar preconceitos existentes sobre padrões de beleza¹⁴. Em específico, o Beauty.AI é examinado como um exemplo de como algoritmos de aprendizado de máquina e redes neurais profundas podem refletir e amplificar vieses socioculturais em representações de imagens corporais.

13 Ibid.

14 Grifos nossos.

A pesquisa de Ding (2023) revela que ferramentas como o Beauty.AI, ao serem treinadas em conjuntos de dados que contêm preferências estéticas historicamente enviesadas, tendem a reproduzir esses preconceitos. Isso é especialmente evidente na preferência por certos tipos de corpos e características faciais frequentemente considerados ideais pela mídia. A análise de capas da revista Vogue, por exemplo, demonstra que os padrões de beleza mudaram ao longo do tempo, mas a preferência por modelos magras e de pele clara persiste, influenciando a percepção de autoimagem e autoestima das pessoas.

Ding (2023) discute a influência da moda e dos padrões de beleza ao longo da história, destacando como a imagem corporal ideal é definida por normas culturais e sociais que variam ao longo do tempo:

A imagem corporal é um tema amplo que pode ser visto da cabeça à cintura e do quadril aos pés, do ponto de vista da moda. Inclui cabelo, corpo, roupas, rostos, pele, cosméticos de beleza e qualquer outra coisa que possa descrever a aparência de uma pessoa (McDowell, 2013 *apud* Ding, 2023, p. 4).

No contexto da inteligência artificial, Ding (2023) critica especificamente o Beauty.AI, um algoritmo desenvolvido para avaliar a beleza humana. O autor argumenta que esses sistemas estão imbuídos de vieses que refletem e reforçam padrões de beleza ocidentais, freqüentemente excluindo diversidade e promovendo estereótipos prejudiciais. Apesar das boas intenções por trás do desenvolvimento de tais tecnologias, a implementação de IA na avaliação da beleza pode resultar em discriminação e perpetuação de normas restritivas de beleza.

Para mitigar esses vieses, Ding (2023) sugere a aplicação de uma abordagem pós-humana no design e na utilização dessas ferramentas de IA. Isso implica a criação de visuais de moda mais inclusivos que abracem a diversidade corporal, utilizando prompts textuais e outras técnicas para orientar os algoritmos de geração de arte a produzir imagens que desafiem os padrões tradicionais de beleza. A visualização de dados é utilizada como método para identificar padrões de viés e informar o desenvolvimento de práticas de design que promovam a inclusão e a representação justa em outputs gerados por IA.

É importante considerar que a questão dos vieses em IA não é exclusiva do Beauty.AI. Pesquisadores como Noble (2018) destacam que algoritmos de busca podem perpetuar estereótipos raciais e de gênero, enquanto O’Neil (2016) argumenta que muitos modelos preditivos usados em áreas como educação e justiça criminal estão profundamente enviesados. Essas críticas sugerem uma necessidade urgente de reavaliar como os conjuntos de dados são selecionados e utilizados no treinamento de sistemas de IA para garantir que eles não perpetuem discriminação e desigualdade existentes.

A inserção cada vez mais profunda da Inteligência Artificial na sociedade traz consigo um espectro de desafios éticos e questões legais intrincadas. À medida que a IA se imiscui em nossas vidas, é inevitável questionar o impacto que ela acarreta. Entre as inquietações prementes, despontam preocupações sobre o emprego, à medida que a automação ameaça determinadas ocupações, levando à necessidade de políticas de requalificação. É válido salientar que, nesse contexto, os mais vulneráveis estão sendo atingidos, e sua regulação necessita ser realizada com urgência, conforme defende o Conselho de Comunicação Social (CSS) do Congresso Nacional e o Projeto de Lei (PL) 2.338/2023¹⁵.

15 Regulação de inteligência artificial é defendida no Conselho de Comunicação. Disponível em: <https://www12.senado.leg.br/noticias/materias/2023/06/05/regulacao-de-inteligencia-artificial-e-defendida-no-conselho-de-comunicacao>. Acesso em: 20 ago. 2023.

Essa reflexão levanta indagações pertinentes acerca do equilíbrio entre eficiência e a preservação da justiça individualizada, valores intrínsecos aos princípios fundamentais do nosso sistema jurídico moderno. Em um cenário em que a inteligência artificial começa a se entrelaçar com a esfera jurídica, o diálogo sobre como manter essa harmonia entre avanço tecnológico, consideração humana e mitigação de vieses preconceituosos se torna especialmente crucial. A coleta e o uso maciço de dados para treinar algoritmos e embasar decisões automatizadas suscitam questionamentos sobre o controle e a propriedade dessas informações, dando lugar a debates éticos sobre vigilância e manipulação de dados pessoais.

Sabemos que a busca pela celeridade processual deve ser uma necessidade primária no sistema jurídico, pois encontra-se respaldada na Emenda nº 45/2004¹⁶, que incorporou ao texto da Constituição o inciso LXXVIII, assegurando a todos, no âmbito judicial e administrativo, a razoável duração do processo e os meios que garantam a celeridade de sua tramitação. No entanto, ao considerarmos essa abordagem, é importante refletir sobre um possível aspecto negativo: a potencial perda da dimensão humana na análise das complexidades legais e na interpretação das situações individuais. Afinal, por mais que essa abordagem pudesse acelerar o andamento dos processos e impulsionar a agilidade da justiça, há um sentimento compartilhado por muitos de que a empatia, a compreensão contextual e o discernimento humano desempenham um papel vital no sistema jurídico.

Ainda que a Inteligência Artificial seja uma ferramenta poderosa para aumentar a eficiência dos processos jurídicos, é importante destacar que ela não substituirá a percepção e o julgamento humano. Como afirma Dreyfus (1992), em sua importante obra “*What Computers Still Can't Do: A Critique of Artificial Reason*”, para ele a habilidade humana de tomar decisões com base em intuição, experiência e percepção contextual é algo que a IA, por mais avançada que seja, não pode replicar completamente. O elemento humano, especialmente em áreas que envolvem questões morais e éticas, continua sendo essencial para garantir uma justiça que vá além da simples aplicação de regras automatizadas.

A privacidade figura como uma preocupação central no cenário mundial. No âmbito brasileiro, o Conselho Nacional de Justiça, ao promulgar a Resolução nº 332 de 2020¹⁷, que versa sobre os princípios de “ética, transparência e governança na concepção e aplicação de sistemas de inteligência artificial no âmbito do Poder Judiciário e estabelece outras diretrizes”, demonstra uma afinidade notável com as orientações delineadas na carta emitida pela Comissão para a Eficácia da Justiça na Europa.

A medida em que a Inteligência Artificial (IA) se enraíza cada vez mais na nossa sociedade, um turbilhão de questões éticas e legais se revela. Essa fusão entre tecnologia e vida cotidiana nos convida a repensar profundamente seu impacto.

Um dilema iminente para uma próxima pesquisa é como a automação alimentada pela IA está alterando nosso mercado de trabalho. À medida que certos empregos são automatizados, surge a urgência de adotar políticas para requalificar os trabalhadores afetados. No entanto, precisamos estar alertas, pois os mais vulneráveis podem sofrer desproporcionalmente. O Con-

16 ROBERTO DE GOUVÉA MEDINA, P. Os valores celeridade processual e segurança jurídica no projeto de novo Código de Processo Civil. [s. l: s. n.]. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/242949/000939995.pdf?sequence=3&isAllowed=y#:~:text=c%C3%ADp%C3%ADo%20constitucional..> Acesso em: 20 ago. 2023.

17 Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 20 ago. 2023.

selho de Comunicação Social (CSS) do Congresso Nacional e o Projeto de Lei (PL) 2.338/2023 chamam a atenção para a necessidade de agir.

No âmbito jurídico, uma dança delicada entre eficiência e justiça individual se desenha. A presença da IA no sistema judicial nos desafia a equilibrar avanços tecnológicos, considerações humanas e a minimização de qualquer parcialidade dos algoritmos. O uso em larga escala de dados para treinar algoritmos e guiar decisões automatizadas levanta questões críticas sobre quem controla esses dados e como eles são usados, desencadeando debates éticos sobre privacidade e manipulação.

Em relação à rapidez nos processos judiciais, buscada com a Emenda nº 45/2004, surge uma preocupação: será que isso pode diminuir a empatia e compreensão humanas na análise de questões legais complexas? A conveniência dos processos automatizados pode colidir com a importância da sensibilidade e compreensão humana.

A privacidade, uma preocupação global, também toca o Brasil. A Resolução nº 332 de 2020 do Conselho Nacional de Justiça vai de encontro aos princípios de “ética, transparência e governança na criação e uso de sistemas de IA no Poder Judiciário”, seguindo orientações similares da Comissão para a Eficácia da Justiça na Europa.

A habilidade da IA de analisar vastos volumes de dados jurídicos e identificar padrões ocultos abre novas portas para a prática jurídica. Contudo, essa adoção requer cuidado para evitar vieses nos algoritmos, garantindo justiça. Nossa análise buscou explorar como a IA molda nossa sociedade, focando nas questões éticas e jurídicas que surgem nesse terreno em constante evolução. Buscamos entender como podemos navegar nesse território permeado pela IA, fomentando um crescimento responsável e equitativo.

A IA já se infiltrou veementemente em nossa sociedade, em nosso dia a dia, gerando debates desde mudanças e adaptações em nossos trabalhos, até desafios legais. Encontrar soluções que aproveitem os benefícios da IA, ao mesmo tempo que protegem nossos valores fundamentais, é um desafio complexo e urgente para os governos, instituições e a sociedade como um todo.

4. CONSIDERAÇÕES FINAIS

As considerações discutidas nesta pesquisa, reverberam um esforço mais amplo na comunidade de pesquisa para abordar e mitigar os impactos negativos da IA enviesada. Essa força científica de trabalho é essencial para promover um uso mais ético e inclusivo da tecnologia, que reconheça e respeite a diversidade humana em todas as suas formas.

A conclusão deste artigo abrange uma reflexão abrangente sobre os desafios éticos e jurídicos inerentes à Inteligência Artificial (IA) e propõe soluções fundamentadas para mitigar os vieses algorítmicos e promover uma justiça mais equitativa. Ao longo do texto, foram discutidos diversos aspectos críticos, desde a evolução histórica da IA até os casos contemporâneos que destacam os perigos dos vieses embutidos em algoritmos.

A introdução delineou a trajetória da IA, desde a concepção da máquina de Turing até os marcos históricos como o primeiro chatbot e a adoção da IA na medicina. Esses avanços tec-

nológicos trouxeram benefícios inegáveis, como a eficiência na análise de dados e a melhoria na prestação de serviços jurídicos e de saúde. Contudo, também evidenciaram problemas graves, como o incidente do chatbot Tay da Microsoft, que destacou os riscos de exposições algorítmicas a dados não moderados, resultando na geração de conteúdos ofensivos e racistas.

O desenvolvimento do artigo aprofundou-se nos desafios éticos e morais, enfatizando a responsabilidade e a transparência como pilares essenciais para a implementação justa da IA. A questão da responsabilidade foi explorada em detalhes, ressaltando a necessidade de um arcabouço legal que clarifique quem é responsável pelas decisões tomadas por sistemas de IA. A falta de transparência, frequentemente descrita como a “caixa preta algorítmica”, foi apontada como um obstáculo significativo à justiça, dificultando a contestação de decisões injustas e a garantia de um processo constitucional devido.

Os exemplos práticos, como o uso de algoritmos no sistema jurídico, ilustraram como a opacidade dos métodos de cálculo pode prejudicar indivíduos, mesmo aqueles com baixo risco de reincidência. A crítica de Nunes e Marques (2018) destacou a impossibilidade de defesa contra índices cuja metodologia é desconhecida, sublinhando a necessidade de maior clareza e auditabilidade nos sistemas algorítmicos.

A solução proposta envolve a implementação de mecanismos de auditoria contínua e revisão dos sistemas de IA, assegurando que estes evoluam para eliminar vieses e promover a justiça. A transparência deve ser promovida por meio de regulamentos que exijam explicações claras e acessíveis sobre o funcionamento dos algoritmos e os dados utilizados. A colaboração interdisciplinar é crucial para desenvolver soluções que garantam uma sociedade mais justa e igualitária, onde a tecnologia seja um instrumento de progresso para todos.

Ademais, a educação e a sensibilização sobre os vieses algorítmicos são essenciais para que desenvolvedores e usuários compreendam as implicações éticas de suas decisões. A transformação digital deve ser guiada por princípios éticos que garantam inclusão e equidade, prevenindo e corrigindo vieses que possam surgir. Como enfatizado por Cortiz (2021), a complexidade do aprendizado profundo requer um esforço adicional para tornar os algoritmos comprehensíveis e auditáveis, assegurando que as decisões sejam justas e transparentes.

Em conclusão, a mitigação dos vieses algorítmicos e a promoção de uma IA mais ética e justa dependem de um compromisso coletivo com a transparência, a responsabilidade e a inclusão. A implementação de regulamentações robustas, a educação contínua e a colaboração interdisciplinar são passos fundamentais para garantir que a tecnologia beneficie a todos de maneira equitativa. Assim, a sociedade pode avançar de forma harmoniosa, equilibrando a inovação tecnológica com os valores éticos essenciais para um futuro mais justo e inclusivo.

REFERÊNCIAS

BARDIN, L. *Análise de conteúdo*. Lisboa: Edições, 70, 225, 1977.

BRASIL. Supremo Tribunal Federal. STF aprova criação da Assessoria de Inteligência Artificial. *STF Notícias*, Brasília, 2023. Disponível em: <https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=471331&ori=1>. Acesso em: 19 out. 2023.

CONSELHO NACIONAL DE JUSTIÇA. *Codex atinge mais de 100 milhões de processos judiciais armazenados.* 2023. Disponível em: <https://www.cnj.jus.br/codex-atinge-mais-de-100-milhoes-de-processos-judiciais-armazenados/>. Acesso em: 19 ago. 2023.

CONSELHO NACIONAL DE JUSTIÇA. Ato nº 3429. 2023. Disponível em: <https://atos.cnj.jus.br/atos/detalhar/3429>. Acesso em: 20 ago. 2023.

CONSELHO DA EUROPA. *Carta ética: traduzida para o português revista.* 2018. Disponível em: <https://rm.coe.int/carta-etica-traduzida-para-portugues-revista/168093b7e0>. Acesso em: 12 maio 2024.

CORTIZ, Diogo. Inteligência artificial: conceitos fundamentais. In: VAINZOF, Rony; GUTIERREZ, Adriei. *Inteligência artificial: sociedade, economia e Estado.* São Paulo: Thomson Reuters, 2021. p. 45-60.

DING, Y. *Desconstruindo a beleza: utilizando inteligência artificial para destacar vieses na indústria da moda.* 2023. Dissertação (Mestrado em Design de Informação e Visualização) - Northeastern University, Boston, 2023

DREYFUS, H. L. *What Computers Still Can't Do: A Critique of Artificial Reason.* Cambridge: MIT Press, 1992.

FREITAS, C. O. A.; CARNEIRO, J. V. V. ; GLASMEYER, R. J. S. . *O Tratamento de Dados Pessoais e Sensíveis frente ao Contact Tracing durante o COVID-19.* Boletim do Grupo de Estudos de Direito Autoral e Industrial: GEDAI/UFPR, Curitiba, p. 1 – 20, 07 jun. 2020. Disponível em: <https://www.gedai.com.br/o-tratamento-de-dados-pessoais-e-sensiveis-frente-ao-contact-tracing-durante-o-covid-19/>

FREITAS, C. O. A.; PAMPLONA, D.A. Cooperação entre Estados Totalitários e Corporações: O uso da segmentação de dados e profiling para violação de direitos humanos. In: RUARO, Regina Linden; MAÑAS, José Luis Piñar; MOLINARO, Carlos Alberto. (org.). *Privacidade e proteção de dados pessoais na sociedade digital.* Porto Alegre: Editora Fi, 2017. v. 1. p. 119-144.

GIL, A. C. *Métodos e técnicas de pesquisa social.* 6. ed. São Paulo: Atlas, 2002.

GREENFIELD, A. *Everyware: A era emergente da computação ubíqua.* AIGA: New Riders, 2006.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Fundamentos de metodologia científica.* 8. ed. São Paulo: Atlas, 2017. ISBN 978-85-970-1076-3.

MENDES, Cleiton. Robôs no tribunal: o papel da inteligência artificial no Judiciário. *Conjur*, Brasília, 15 dez. 2023. Disponível em: <https://www.conjur.com.br/2023-dez-15/robos-no-tribunal-papel-da-inteligencia-artificial-no-judiciario/>. Acesso em: 18 abr. 2024.

MENDES, Cleiton. O uso da IA no sistema de justiça é um dos grandes desafios do século. *Conjur*, Brasília, 5 abr. 2024. Disponível em: <https://www.conjur.com.br/2024-abr-05/uso-da-ia-no-sistema-de-justica-e-um-dos-grandess-desafios-do-seculo/>. Acesso em: 18 abr. 2024.

MENDES, Cleiton. Burocrático, ineficiente e o Estado é o maior litigante do Brasil. *Conjur*, Brasília, 11 mar. 2023. Disponível em: <https://www.conjur.com.br/2023-mar-11/burocratico-ineficiente-estado-maior-litigante-brasil/>. Acesso em: 18 abr. 2024.

NOBLE, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism.* New York University Press, 2018. Acesso em: 11 abr. 2024.

NUNES, Dierle; MORATO, Otávio. A explicabilidade da inteligência artificial e o devido processo tecnológico. *Conjur*, São Paulo, 7 jul. 2021. Disponível em: <https://www.conjur.com.br/2021-jul-07/opiniao-explicabilidade-ia-devidoprocesso-tecnologico>. Acesso em: 21 maio 2024.

NUNES, Dierle; MARQUES, Ana Luiza. Inteligência artificial e direito processual: vieses algorítmicos e os riscos de atribuição de função decisória às máquinas. *Revista de Processo*, v. 285, p. 421-447, nov. 2018. Disponível em: https://www.academia.edu/37764508/INTELIG%C3%8ANCIA_ARTIFICIAL_E_DIREITO_PROCESSUAL_VIESSES_ALGOR%C3%80DTMICOS_E_OS_RISCOS_DE_ATRIBUI%C3%87%C3%83O_DE_FUN%C3%87%C3%83O_DECIS%C3%93RA_%C3%80S_M%C3%81QUINAS_Artificial_intelligence_and_procedural_law_algorithmic_bias_and_the_risks_of_assignment_of_decision_making_function_to_machines. Acesso em: 21 maio 2024.

O'NEIL, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group, 2016.

Antonio Felipe de Sousa e Renata Albuquerque Lima

ROBERTO DE GOUVÉA MEDINA, P. *Os valores celeridade processual e segurança jurídica no projeto de novo Código de Processo Civil.* [s. l.: s. n.]. Disponível em: <https://www2.senado.leg.br/bdsf/bitstream/handle/id/242949/000939995.pdf?sequence=3&isAllowed=y#:~:text=c%C3%ADpio%20constitucional..> Acesso em: 25 ago. 2023.

SENADO FEDERAL. *Regulação de inteligência artificial é defendida no Conselho de Comunicação.* 2023. Disponível em: <https://www12.senado.leg.br/noticias/materias/2023/06/05/regulacao-de-inteligencia-artificial-e-defendida-no-conselho-de-comunicacao>. Acesso em: 20 ago. 2023.

SENADO FEDERAL. *Conselho de Comunicação Social discute uso da inteligência artificial nos meios de comunicação.* 2023. Disponível em: <https://www12.senado.leg.br/noticias/materias/2023/06/05/regulacao-de-inteligencia-artificial-e-defendida-no-conselho-de-comunicacao>. Acesso em: 21 maio 2024.

TURING, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem". *Proceedings of the London Mathematical Society*, v. 42, n. 2, p. 230-265, 1936.

Dados do processo editorial

- Recebido em: 11/09/2024
- Controle preliminar e verificação de plágio: 13/09/2024
- Avaliação 1: 27/09/2024
- Avaliação 2: 08/01/2025
- Decisão editorial preliminar: 08/01/2025
- Retorno rodada de correções: 27/01/2025
- Decisão editorial/aprovado: 16/03/2025

Equipe editorial envolvida

- Editor-chefe: 1 (SHZF)
- Editor-assistente: 1 (ASR)
- Revisores: 2