

**II CONGRESSO INTERNACIONAL DE
DIREITO, POLÍTICAS PÚBLICAS,
TECNOLOGIA E INTERNET**

**ACESSO À JUSTIÇA, INTELIGÊNCIA ARTIFICIAL E
TECNOLOGIAS DO PROCESSO JUDICIAL**

A174

Acesso à justiça, inteligência artificial e tecnologias do processo judicial [Recurso eletrônico on-line] organização II Congresso Internacional de Direito, Políticas Públicas, Tecnologia e Internet: Faculdade de Direito de Franca – Franca;

Coordenadores: Elizabete Cristiane de Oliveira Futami, Angela Issa Haonat e Caio Augusto Souza Lara – Franca: Faculdade de Direito de Franca, 2024.

Inclui bibliografia

ISBN: 978-65-5274-023-6

Modo de acesso: www.conpedi.org.br em publicações

Tema: Regulação do Ciberespaço.

1. Acesso à Justiça. 2. Inteligência Artificial. 3. Processo Judicial Eletrônico. 4. Políticas Públicas de Desenvolvimento. 5. Efetividade do Direito. I. II Congresso Internacional de Direito, Políticas Públicas, Tecnologia e Internet (1:2024 : Franca, SP).

CDU: 34

II CONGRESSO INTERNACIONAL DE DIREITO, POLÍTICAS PÚBLICAS, TECNOLOGIA E INTERNET

ACESSO À JUSTIÇA, INTELIGÊNCIA ARTIFICIAL E TECNOLOGIAS DO PROCESSO JUDICIAL

Apresentação

Entre os dias 27 e 30 de agosto de 2024, a Faculdade de Direito de Franca recebeu o Congresso Internacional de Direito, Políticas Públicas, Tecnologia e Internet. O evento reuniu acadêmicos, profissionais, pesquisadores e estudantes, promovendo o debate interdisciplinar sobre o impacto das inovações tecnológicas no campo jurídico e nas políticas públicas. A programação envolveu Grupos de Trabalho (GTs) organizados para aprofundar temas específicos, abordando desde o acesso à justiça até as complexidades da regulação tecnológica, com ênfase na adaptação do sistema jurídico aos avanços da inteligência artificial e da automação.

O GT 1 – Acesso à Justiça, Inteligência Artificial e Tecnologias do Processo Judicial focou na relação entre o acesso à justiça e o uso de tecnologias avançadas no processo judicial, abordando as múltiplas formas de acesso — formal, material e pelos direitos fundamentais. Discutiu-se como o uso da inteligência artificial pode moldar o futuro do judiciário, com debates sobre as ações do Conselho Nacional de Justiça para regulamentar o uso dessas ferramentas, além de questões éticas e de inovação no processo judicial eletrônico. As contribuições deste GT exploram o papel da jurimetria e da juscibernética na construção de um sistema de justiça mais acessível e eficiente, e propõem novas perspectivas para o futuro do direito na era digital.

FAZENDO MÁQUINAS TOMAREM DECISÕES: UMA ANÁLISE SOBRE PERPETUAÇÃO DE VIESES, TRANSPARÊNCIA E DISCRIMINAÇÃO NA IA

MAKING MACHINES TAKE DECISIONS: AN ANALYSIS OF THE PERPETUATION OF BIAS, TRANSPARENCY AND DISCRIMINATION IN AI

**Pedro Henrique do Prado Haram Colucci
Sergio Nojiri**

Resumo

O presente trabalho investiga os desafios éticos na utilização da Inteligência Artificial (IA) na tomada de decisões, especialmente no campo jurídico. Tem-se como objetivo, a partir de uma revisão de literatura, destacar os desafios éticos e práticos associados à utilização dessas tecnologias. Partindo da hipótese de que a transposição da agência de decisão para as máquinas alimentadas por algoritmos de bancos de dados não mitigam ou blindam os resultados de serem maculados por vieses, pelo contrário, descortina-se uma nova dimensão de desafios para pensar a interdisciplinaridade entre direito e psicologia cognitiva nos processos de tomada de decisão automatizada.

Palavras-chave: Discriminação algorítmica, Vieses, Tomada de decisão, Psicologia cognitiva

Abstract/Resumen/Résumé

This paper investigates the ethical challenges of using Artificial Intelligence (AI) in decision-making, especially in the legal field. Based on a literature review, the aim is to highlight the ethical and practical challenges associated with the use of these technologies. Based on the hypothesis that the transposition of the decision-making agency to machines fed by database algorithms does not mitigate or shield the results from being tainted by biases; on the contrary, it reveals a new dimension of challenges for thinking about the interdisciplinarity between law and cognitive psychology in automated decision-making processes.

Keywords/Palabras-claves/Mots-clés: Algorithmic discrimination, Biases, Decision making, Cognitive psychology

INTRODUÇÃO

O conceito de Inteligência Artificial (IA) é frequentemente empregado enquanto resultado do desenvolvimento de algoritmos de aprendizagem automática que são treinados para reconhecer padrões e tendências em conjuntos massivos de dados, rotulados como *Big Data*. Em síntese, trata-se da utilização de algoritmos, dados e modelos computacionais para simular aspectos da inteligência humana para determinadas tarefas.

O processo de aprendizagem automática utiliza um extenso repositório de informações usado para construir sua base de conhecimento e aprender os parâmetros ideais de um modelo e reproduzir com precisão resoluções de problemas específicos a partir de resultados conhecidos quando acionada. Experiências e fatos passados são usados como base para instruir a IA a prever resultados futuros (Barabas, 2019).

A diversidade e riqueza do banco de dados fornecido impacta diretamente a capacidade da IA de executar determinadas tarefas com precisão, como reconhecimento facial e policiamento preditivo. O conjunto de dados não surge do nada, todos os dados que alimentam o repertório da IA se originam de contexto sociais e são constituídos e atravessados por questões epistemológicas, visões de mundo, preconceitos e tendências, o que pode fazer com que os sistemas herdem e ampliem involuntariamente essas tendências, isto é, segundo Sartori e Theodorou, a IA é essencialmente um aparato "sociotécnico" (2022, p. 2, tradução nossa).

Com isso posto, o presente trabalho explora os vieses e heurísticas que afetam tanto humanos quanto sistemas de IA. Tem-se como objetivo, a partir de uma revisão de literatura, destacar os desafios éticos e práticos associados à utilização dessas tecnologias. Partindo da hipótese de que a transposição da agência de decisão para as máquinas alimentadas por algoritmos de bancos de dados não mitigam ou blindam os resultados de serem maculados por vieses, pelo contrário, descortina-se uma nova dimensão de desafios para pensar a interdisciplinaridade entre direito e psicologia cognitiva nos processos de tomada de decisão automatizada.

DESENVOLVIMENTO

No campo das tecnologias de reconhecimento facial, Alex Najibi (2020) aponta como solução para mitigar os processos de enviesamento da IA ampliar e diversificar o conjunto de dados utilizado para treinamento dos algoritmos. No entanto, pesquisadores da ciência da computação (Denton; Hutchinson; Mitchell; Gebru; Zaldivar, 2020) já alertaram que quanto maior for o conjunto de dados utilizado, maior será o risco de preconceitos e deturpações incorporados, tendo como base o atual nível de desinformação produzido pelas atualizações do ChatGPT-4.

A aplicação de tecnologias de inteligência artificial no sistema jurídico tinha como promessa revolucionar a maneira como as decisões judiciais são tomadas, oferecendo análises imparciais, rápidas, consistentes e baseadas em dados. No entanto, essa implementação não está isenta de desafios (Morais da Rosa, 2020). Um dos principais pontos de discussão é a presença de vieses e heurísticas tanto em decisões humanas quanto nas tomadas por IA (Fenoll, 2018).

Os psicólogos Mellers, Schwartz e Cooke (1988), em suas pesquisas sobre comportamento judicial, buscam tensionar os complexos aspectos envolvidos na atividade de tomada de decisões humanas. Tradicionalmente, a teoria da escolha racional tem sido o paradigma dominante para entender como os indivíduos fazem escolhas. No entanto, os autores argumentam que a tomada de decisões humanas não se limita à lógica pura e à otimização de utilidade.

Em vez disso, uma série de fatores, incluindo emoções, crenças, contexto e modelos alternativos de escolha, influenciam nossas decisões de maneiras complexas e variadas. Essa perspectiva destaca a importância das normas sociais, éticas e morais na tomada de decisões. Os autores também exploram a abordagem baseada em preferências, que sugere que as escolhas são moldadas por avaliações subjetivas de crenças e valores. Essa perspectiva enfatiza a individualidade da tomada de decisões e como as experiências passadas moldam nossas preferências.

Outro modelo apresentado é a visão de que algumas decisões podem ser automáticas e não requerem o processo de ponderação de alternativas, destacando a importância da intuição e da familiaridade nas escolhas cotidianas. O contexto exerce uma influência significativa sobre a tomada de decisões, isto é, o estímulo do contexto, que inclui o ambiente

em que uma decisão é tomada e as experiências passadas do tomador de decisões, pode moldar as nossas escolhas.

As escolhas são simplificadas por meio de heurísticas, isto é, atalhos cognitivos. Isso destaca a tendência humana de simplificar decisões complexas, muitas vezes em detrimento da racionalidade pura. A tomada de decisão judicial humana é suscetível a uma variedade de vieses psicológicos (Kahneman, Tversky, 1974). Entre eles, o viés de confirmação, onde julgadores tendem a buscar informações que confirmem suas hipóteses prévias. O viés de ancoragem, onde decisões são influenciadas por informações iniciais apresentadas (como uma primeira oferta em negociações), e o viés de representatividade, que leva a decisões baseadas em estereótipos e generalizações, são igualmente preocupantes. Estes vieses podem levar a decisões subjetivas e inconsistentes, mesmo entre juízes experientes e bem treinados.

Alexandre de Moraes da Rosa e Paola Wojciechowski (2021, p. 99) apontam que, no contexto em que máquinas progressivamente começaram a assumir funções sociais de decisão, passou-se a depositar "[...] nas novas tecnologias a esperança de que processos de tomadas de decisão — inclusive judiciais — pudessem ser depurados de vieses e subjetivismo". Essas expectativas se mostraram ingênuas, enquanto vieses psicológicos têm sido amplamente estudados em contextos humanos, sua transferência para algoritmos de IA pode resultar em decisões automatizadas que perpetuam injustiças.

A transferência inadvertida de vieses humanos para sistemas de IA é uma preocupação central. Dados oriundos de contextos socioculturais refletem preconceitos sociais e estruturais, introduzindo vieses nos algoritmos de aprendizado de máquina, como dados criminais que refletem disparidades raciais devido a políticas criminais enviesadas. Os sistemas de IA podem perpetuar e até amplificar essas disparidades, mesmo sob o disfarce de objetividade, quando o *data set* empregado no treinamento da IA já nasce enviesado (Moraes da Rosa; Wojciechowski, 2021, p. 100).

Tavares (2021) aponta como exemplo a utilização de uma ferramenta para assessorar juízes na aplicação de penas em processos criminais alimentada por uma base de dados de sentenças pretéritas, essa dinâmica pode inadvertidamente perpetuar e amplificar os mesmos preconceitos já contidos em julgadores, incorporando vieses passados na feitura das decisões presentes.

Dados oriundos de decisões humanas calcadas por desigualdades sociais e variáveis como gênero, raça e nacionalidade uma vez introduzidos no na base de dados da IA, o viés pode ser corrigido se a fonte tendenciosa for detectada. No entanto, uma vez presentes deturpações no sistema, o resultado viciado já está no mundo.

CONSIDERAÇÕES FINAIS

Este trabalho explorou os desafios associados à utilização da Inteligência Artificial, especialmente no contexto da tomada de decisões. Ao revisar a literatura, ficou claro que a promessa inicial de imparcialidade e objetividade das decisões tomadas por IA enfrenta sérios obstáculos. A transferência inadvertida de vieses humanos para sistemas de IA, através dos dados que alimentam seus algoritmos, revelou-se uma questão central.

Os sistemas de IA são alimentados por grandes conjuntos de dados que refletem não apenas informações, mas também preconceitos sociais e estruturais presentes em contextos históricos e culturais. Esses dados podem introduzir e amplificar vieses, reproduzindo disparidades raciais, de gênero e outras formas de discriminação, mesmo sob a fachada de neutralidade e precisão técnica.

No campo jurídico, onde a IA passa a ser pensada a partir da promessa de melhorar a coesão e a imparcialidade das decisões judiciais, os desafios restam particularmente evidentes. Os vieses psicológicos que afetam os julgadores humanos também foram identificados nos sistemas de IA, influenciando decisões automatizadas de maneira potencialmente injusta.

Portanto, é necessário reconhecer que a simples automatização de processos decisórios não garante a eliminação de vieses; pelo contrário, pode perpetuá-los de maneiras novas e insuspeitas. A detecção e mitigação desses vieses exigem uma abordagem interdisciplinar que integre conhecimentos de direito, ética, psicologia cognitiva e ciência de dados. A conscientização e a atuação contraintuitiva nesse cenário é o primeiro passo para desenvolver políticas e práticas regulatórias que garantam que a IA seja implementada de maneira ética e responsável.

REFERÊNCIAS

BARABAS, Chelsea. Beyond Bias: "Ethical AI" in Criminal Law. In: **The Oxford Handbook of Ethics of AI**. Oxford: Oxford University Press, 2019. pp. 737-753.

DENTON, Remi; HUTCHINSON, Ben; MITCHELL, Margaret; GEBRU, Timnit; ZALDIVAR, Andrew. Image counterfactual sensitivity analysis for detecting unintended bias.

In: **CVPR 2019 Workshop on Fairness Accountability Transparency and Ethics in Computer Vision**. 2020. pp. 1-12.

FENOLL, Jordi Nieva. **Inteligencia artificial y proceso judicial**. Marcial Pons: Madrid, Barcelona, Buenos Aires, São Paulo, 2018.

MELLERS, B. A.; SCHWARTZ, A.; COOKE, A. D. J. Judgment and decision making. **Annual Review of Psychology**, v. 49, p. 447–477, 1998.

MORAIS DA ROSA, Alexandre. **Ensinando um Robô a Julgar**: pragmática, discricionariiedade, heurísticas e vieses no uso de aprendizado de máquina no Judiciário. Florianópolis: EMais, 2020.

NAJIBI, Alex. **Racial Discrimination in Face Recognition Technology**. Harvard University Blog, Science Policy and Social Justice Edition. 2020. Disponível em: <http://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/>.

SARTORI, Laura; THEODOROU, Andreas. A sociotechnical perspective for the future of AI: narratives, inequalities, and human control. In: **Ethics Inf Technol**. v. 24, n. 4, pp. 1-11. 2022.

TAVARES, João Paulo Lordelo Guimarães. Vieses Implícitos e técnicas de automação decisória: riscos e benefícios. In: **Civil Procedure Review**, v. 12, n. 1, pp. 105-132. jan.-abr. 2021.

TVERSY, Amos; KAHNEMAN, Daniel. Judgment under Uncertainty: Heuristics and biases. **Science**, v. 185, p. 1.124-1.131, 1974.

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