

DESAFIOS ÉTICOS AO USO DA INTELIGÊNCIA ARTIFICIAL NO SISTEMA DE JUSTIÇA CRIMINAL

ETHICAL ISSUES TO THE USE OF ARTIFICIAL INTELLIGENCE IN THE CRIMINAL JUSTICE SYSTEM

Anderson de Andrade Bichara¹  

Comunidade de Polícias de Américas, Ameripol, Colômbia
andersonbichara@gmail.com

Fabiana Amaro de Brito²  

Universidad Internacional Iberoamericana,
Unini, México
fabiana.amaro@doctorado.unini.edu.mx

DOI: <https://doi.org/10.5281/zenodo.13630977>

Resumo: O presente artigo examina os desafios éticos envolvidos no uso da inteligência artificial (IA) no sistema de justiça criminal. Inicialmente, a IA foi celebrada por suas promessas de aprimorar a elucidação de crimes e a criminologia preditiva, mas casos de aplicação equivocada e reforço de preconceitos levantam questões importantes. Para este artigo, utilizamos uma revisão bibliográfica qualitativa, analisando casos práticos e estudos acadêmicos relevantes para destacar os desafios e as implicações éticas do uso de IA no sistema de justiça criminal. A bibliografia foi escolhida por seu rigor acadêmico e sua relevância atual, abordando tanto os sucessos quanto os fracassos da IA, além de fornecer uma visão crítica sobre os vieses e as consequências sociais do uso dessas tecnologias. A seleção das fontes visou garantir uma compreensão abrangente e equilibrada do tema, com foco na justiça e equidade.

Palavras-chave: sistemas inteligentes; sistema judiciário; deontologia; vieses; justiça social; criminologia preditiva.

Abstract: This article examines the ethical challenges involved in the use of artificial intelligence (AI) in the criminal justice system. Initially, AI was celebrated for its potential to enhance crime elucidation and predictive criminology, but instances of misapplication and reinforcement of biases raise significant concerns. A qualitative literature review was conducted, analyzing practical cases and relevant academic studies to highlight the challenges and ethical implications of AI usage in the criminal justice system. The selected bibliography was chosen for its academic rigor and current relevance, addressing both the successes and failures of AI, and providing a critical view on biases and the social consequences of these technologies. The aim is to offer a comprehensive and balanced understanding of the subject, with a focus on justice and equity.

Keywords: intelligent automation; judicial system; deontology; biases; social justice; predictive criminology.

1. Introdução

Nos últimos anos, a inteligência artificial (IA) emergiu como uma ferramenta promissora no campo da criminologia e na elucidação de crimes. A capacidade de analisar grandes volumes de dados e identificar padrões complexos permite que a IA ofereça novas perspectivas e soluções para problemas antigos. A aplicação da

IA em segurança pública inclui desde a previsão de crimes até a análise de vídeos de segurança, o que aumenta potencialmente a eficiência e a eficácia das forças de segurança.

A promessa inicial da IA era a de uma ferramenta imparcial e objetiva, capaz de mitigar erros humanos e preconceitos. Algoritmos avançados de reconhecimento facial e análise preditiva de crimes foram desenvolvidos para auxiliar na identificação de suspeitos e

¹ Mestre em Criminologia Aplicada e Investigación Policial (UCAV-Espanha, 2023). Especialização em Gestão de Órgãos de Segurança Pública (IUPPERJ-EPPG). Especialização em Direito e Processo do Trabalho (UGF). Bacharel em Direito (UFES). Delegado do Secretário-Executivo da AMERIPOL – Comunidade de Polícias de Américas. Link Lattes: <http://lattes.cnpq.br/3648943527323618>. ORCID: <https://orcid.org/0000-0002-7995-6720>.

² Doutoranda em Psicologia – Psicologia Forense pela Unini. Mestre em Ciência Policiais (Instituto Superior de Ciências Policiais e Segurança Interna – Lisboa/Portugal). Pós-graduação em filosofia e sociologia. Pós-graduação em gestão da segurança. Pós-graduação em segurança pública, cultura e cidadania. Oficial da PMERJ. Curso de Formação de Oficiais pela Academia de Polícia Militar D. João VI (2003). Link Lattes: <http://lattes.cnpq.br/1473899455667291>. ORCID: <https://orcid.org/0000-0002-7179-7155>.

na alocação de recursos policiais de forma mais eficiente. Esses avanços indicavam um futuro onde a tecnologia poderia ajudar a criar um sistema de justiça mais justo e eficiente.

No entanto a realidade mostrou-se mais complexa. Ao invés de eliminar preconceitos, a IA frequentemente os reforça, reproduzindo discriminações existentes no sistema de justiça criminal. As bases de dados utilizadas para treinar algoritmos muitas vezes contêm vieses históricos, que acabam sendo perpetuados pelas tecnologias de IA. Isso levanta importantes questões éticas sobre a implementação dessas tecnologias em um campo tão sensível quanto o da justiça criminal (**Barabas**, 2020).

Ademais, o entusiasmo inicial com a IA deve ser moderado por uma análise crítica dos seus impactos sociais e éticos. Como apontado por autores como **Noble** (2018), algoritmos de IA podem perpetuar desigualdades sociais e preconceitos se não forem cuidadosamente desenhados e monitorados. A falta de transparência nos processos algorítmicos e a complexidade técnica podem dificultar a identificação e a correção de vieses. É fundamental que os desenvolvedores de IA e os responsáveis pela sua implementação estejam cientes desses riscos e trabalhem ativamente para mitigá-los.

Além disso, a implementação da IA no sistema de justiça criminal deve ser acompanhada de uma rigorosa supervisão ética e legal. Isso inclui a criação de diretrizes claras sobre a responsabilidade e a *accountability* no uso de tecnologias de IA, garantindo que os direitos dos indivíduos sejam protegidos. Como destacado por **Eubanks** (2018), a transparência e a prestação de contas são essenciais para evitar abusos e garantir que a IA seja utilizada de maneira justa e equitativa.

2. Utilização da IA

A aplicação da IA na segurança pública é repleta de méritos. Em várias localidades, a tecnologia tem sido utilizada com sucesso para aumentar a eficiência das operações policiais e melhorar a segurança da comunidade. Por exemplo, o uso de algoritmos de reconhecimento facial tem permitido a identificação rápida de suspeitos em grandes eventos públicos, auxiliando na prevenção de crimes.

Outra área de sucesso é a análise preditiva de crimes, onde a IA é usada para identificar áreas de alto risco e prever possíveis atividades criminosas. Essa abordagem tem permitido uma alocação mais estratégica de recursos policiais, resultando em uma redução significativa da criminalidade em algumas regiões. Além disso, tecnologias de detecção de tiros baseadas em IA têm sido implementadas em várias cidades, permitindo respostas mais rápidas e eficazes a incidentes violentos.

A análise de vídeo assistida por IA também tem mostrado resultados promissores. Algoritmos avançados são capazes de analisar horas de gravações de segurança em questão de minutos, identificando comportamentos suspeitos e fornecendo pistas valiosas para investigações criminais. Essas tecnologias não só economizam tempo e recursos, mas também aumentam a precisão das investigações.

Outra aplicação de sucesso da IA pode ser vista no uso de sistemas de monitoramento de dados para prever comportamentos criminosos. Por exemplo, em algumas cidades, algoritmos têm sido usados para analisar padrões de chamadas de emergência e identificar áreas onde crimes são mais prováveis de ocorrer. Esse tipo de análise preditiva permite que as forças de segurança se concentrem em áreas específicas e adotem medidas preventivas antes que os crimes ocorram.

Adicionalmente, a IA tem sido utilizada para melhorar a eficiência dos processos judiciais. Sistemas de IA são capazes de analisar

grandes volumes de documentos legais e extrair informações relevantes, auxiliando juízes e advogados na tomada de decisões mais informadas. Essa automação dos processos pode reduzir significativamente o tempo e os custos associados aos litígios. Esses exemplos demonstram como a IA pode ser uma ferramenta poderosa para aumentar a eficiência e a eficácia do sistema de justiça criminal.

Finalmente, tecnologias de IA têm sido empregadas para monitorar comportamentos de indivíduos em liberdade condicional ou em regime semiaberto. Algoritmos podem analisar dados de localização e comportamento desses indivíduos para detectar padrões que possam indicar riscos de reincidência. Essas ferramentas permitem uma supervisão mais precisa e podem contribuir para a reintegração mais segura desses indivíduos na sociedade.

3. Uso disfuncional da IA

Apesar dos casos de sucesso, o uso da IA na justiça criminal nem sempre resulta em justiça. Em alguns casos, a aplicação equivocada da tecnologia tem levado a injustiças significativas. Um exemplo notório é o uso de sistemas de reconhecimento facial que, devido a vieses nos dados de treinamento, apresentam altas taxas de erro na identificação de indivíduos de minorias étnicas (**Barabas**, 2020). Esses erros podem levar a prisões injustas e agravar a desconfiança nas forças de segurança.

Além disso, a análise preditiva de crimes, que visa alocar recursos policiais de forma mais eficiente, tem sido criticada por reforçar estigmas e discriminações. Algoritmos que utilizam dados históricos de criminalidade tendem a direcionar mais policiamento para comunidades já marginalizadas, perpetuando ciclos de vigilância e criminalização excessiva (**Barabas**, 2020). Essa prática não só falha em abordar as causas profundas da criminalidade, como também contribui para a deterioração das relações entre a polícia e a comunidade.

Casos de erro de identificação em sistemas de IA também são preocupantes. Erros em algoritmos de reconhecimento facial, por exemplo, podem levar a acusações injustas, como evidenciado por casos em que indivíduos foram falsamente identificados como criminosos com base em correspondências errôneas de IA. Tais incidentes destacam a necessidade de cautela e supervisão humana contínua na aplicação dessas tecnologias.

Além disso, o uso de IA para determinar penas e fianças judiciais tem gerado controvérsias. Em alguns casos, sistemas automatizados foram encontrados reforçando disparidades raciais ao recomendarem penas mais severas para minorias étnicas em comparação com réus brancos em situações semelhantes. Esse problema ocorre porque os algoritmos podem refletir preconceitos existentes nos dados de treinamento, perpetuando injustiças ao invés de mitigá-las (**Angwin et al.**, 2016).

A falta de transparência nos algoritmos de IA também é um problema significativo. Muitas vezes, as decisões tomadas por esses sistemas são difíceis de explicar ou contestar, o que pode dificultar a busca por justiça. Quando as pessoas não entendem como e por que foram tomadas determinadas decisões, a confiança no sistema judicial pode ser comprometida. Isso é especialmente problemático em contextos onde a liberdade individual está em jogo (**Eubanks**, 2018).

Finalmente, a dependência excessiva de IA pode levar à desumanização do processo judicial. Decisões que antes eram tomadas por humanos, considerando nuances e contextos específicos, passam a ser baseadas em modelos matemáticos. Embora a IA possa fornecer suporte valioso, é essencial manter a supervisão humana para garantir que as decisões sejam justas e éticas. A justiça deve ser sempre centrada na humanidade, e a

tecnologia deve ser uma ferramenta para servir, não para substituir, essa premissa.

4. Para além dos vieses

O desafio de lidar com os vieses nos algoritmos de IA é complexo e multifacetado. Algoritmos são frequentemente treinados em bases de dados que refletem as desigualdades e discriminações presentes na sociedade. Como resultado, eles tendem a reproduzir esses vieses, perpetuando preconceitos sistêmicos. Isso é especialmente problemático no contexto da justiça criminal, onde decisões enviesadas podem ter consequências graves para os indivíduos afetados (Barabas, 2020).

A questão dos vieses não se limita apenas aos dados, mas também aos próprios processos de *design* e implementação dos algoritmos. Pesquisas mostram que a falta de diversidade nas equipes de desenvolvimento pode contribuir para a criação de sistemas que não consideram adequadamente as necessidades e circunstâncias de diferentes grupos sociais (Barabas, 2020). Abordar esses desafios requer uma abordagem interdisciplinar que combine *insights* técnicos com uma compreensão profunda das implicações sociais e éticas.

É essencial que desenvolvedores e implementadores de IA adotem práticas rigorosas de auditoria e transparência para identificar e mitigar vieses. Isso inclui o uso de métodos de *fairness-aware machine learning*, que buscam equilibrar a precisão dos algoritmos com a necessidade de justiça e equidade (Barabas, 2020). *Fairness-aware* refere-se a práticas de aprendizado de máquina (*machine learning*) que são explicitamente projetadas para minimizar vieses e garantir que os resultados dos algoritmos sejam justos e imparciais (Barocas; Hardt; Narayanan, 2023). Aprendizado de máquina é um ramo da IA que se concentra no desenvolvimento de algoritmos que permitem que as máquinas aprendam a partir de dados e melhorem seu desempenho ao longo do tempo sem serem explicitamente programadas para cada tarefa (Mitchell, 1997).

A implementação de diretrizes claras e políticas de governança é fundamental para lidar com os vieses em IA. Essas diretrizes devem incluir a revisão contínua dos algoritmos e a coleta de *feedback* de diversos grupos sociais para garantir que as decisões sejam justas e equitativas (Eubanks, 2018). A transparência nos processos de desenvolvimento e a inclusão de múltiplas perspectivas podem ajudar a mitigar os riscos associados aos vieses algorítmicos.

Além disso, é importante considerar o papel da regulação governamental na supervisão do uso da IA no sistema de justiça criminal. Políticas públicas devem ser implementadas para garantir que a IA seja utilizada de forma ética e responsável. Isso inclui a criação de órgãos reguladores independentes que possam monitorar e avaliar o impacto das tecnologias de IA na justiça criminal, assegurando que os direitos dos indivíduos sejam protegidos.

A teoria da justiça de John Rawls (1971) fornece uma estrutura essencial para avaliar o uso ético da IA no sistema de justiça criminal. Segundo o autor, uma sociedade justa é aquela em que as desigualdades são organizadas de forma a beneficiar os menos favorecidos e onde há igualdade de oportunidades. Aplicando esses princípios ao uso de IA, é fundamental garantir que as tecnologias não reforcem desigualdades existentes, mas sim que contribuam para uma distribuição mais equitativa de recursos e oportunidades. Além disso, a teoria da justiça de Rawls (1971) destaca a importância do "véu de ignorância", onde decisões devem ser tomadas sem conhecimento das posições sociais dos indivíduos envolvidos. Isso implica que o desenvolvimento e a implementação de IA no sistema de justiça devem ser orientados por princípios de imparcialidade e equidade, garantindo que todos os indivíduos sejam tratados

de forma justa, independentemente de sua origem social, racial ou econômica.

A ética pré e pós-aristotélica também oferece *insights* valiosos para a discussão sobre IA e justiça. A ética aristotélica, com seu foco na virtude e no bem-estar humano, sugere que as ações devem ser desenvolvidas e utilizadas de maneira a que promovam o florescimento humano e o bem comum (Aristóteles, 2009, p. 110). Se transpusermos esses postulados universais para nossa discussão, temos que o uso da IA requer uma consideração cuidadosa dos impactos sociais e individuais que as novas tecnologias trarão. Uma ponderação inevitável nessa equação diz respeito a considerar sempre o máximo de resultado, em termos de promoção de justiça, com o mínimo de risco, ou seja, com redução da possibilidade de que se cometam injustiças ou se perpetuem estigmas e preconceitos.

Por outro lado, a ética pós-aristotélica, influenciada por pensadores como Kant (2002) e Mill (1998), enfatiza a importância de princípios universais e do utilitarismo na tomada de decisões éticas. Aplicando esses princípios ao contexto da IA, é essencial que as tecnologias sejam desenvolvidas com base em normas éticas rigorosas que garantam a dignidade e os direitos de todos os indivíduos, além de maximizar o bem-estar coletivo.

Bichara e Cascardo Junior (2023) argumentam, no que diz respeito à investigação, que a proibição de viés de confirmação é essencial para garantir a imparcialidade em inquéritos policiais. Eles destacam que o viés de confirmação pode levar a apurações injustas, em que os investigadores põem ênfase nas evidências que confirmem suas hipóteses iniciais, ignorando provas que possam apontar para outra direção, *id est*, que infirmem sua hipótese-criminal. Essa prática não só compromete a integridade das investigações, mas também pode resultar em injustiças significativas, especialmente quando aplicadas a minorias e grupos marginalizados ou quando se levam em consideração pressões midiáticas ou de qualquer outra natureza como móvel da investigação policial.

5. Vieses de um sistema de justiça criminal discriminatório

O sistema de justiça criminal, historicamente marcado por práticas discriminatórias, corre o risco de ter essas injustiças amplificadas pelo uso inadequado de IA. Algoritmos utilizados para prever o risco de reincidência frequentemente penalizam desproporcionalmente indivíduos de minorias étnicas. Esses algoritmos, ao se basearem em dados históricos de prisões e condenações, refletem e perpetuam os preconceitos embutidos nesses dados.

Além disso, a utilização de IA para decisões de fiança e sentenças judiciais levanta preocupações significativas. Sistemas automatizados que recomendam penas ou decisões de liberdade condicional podem introduzir um nível de desumanização no processo judicial, em que decisões são tomadas com base em modelos matemáticos ao invés de uma consideração holística do indivíduo. Essa prática não só compromete a justiça individual, mas também mina a confiança pública no sistema judicial.

A crítica central é que a adoção de IA em um sistema já preconceituoso pode agravar desigualdades existentes, em vez de mitigá-las. A falta de transparência nos algoritmos e a ausência de mecanismos robustos de responsabilização dificultam a identificação e a correção de erros e injustiças. Portanto, a implementação dessas tecnologias deve ser acompanhada de salvaguardas rigorosas para proteger os direitos dos indivíduos e garantir que a justiça seja verdadeiramente servida.

Para mitigar esses riscos, é fundamental que os desenvolvedores de IA trabalhem em estreita colaboração com especialistas em ética e direitos humanos. Essa colaboração pode ajudar a identificar e abordar potenciais fontes de discriminação antes que os algoritmos sejam implementados em larga escala (Barabas, 2020). Além disso,

a criação de comitês de revisão ética pode fornecer uma supervisão contínua, garantindo que as práticas de desenvolvimento e uso da IA estejam alinhadas com os princípios de justiça e equidade.

A educação e o treinamento de profissionais do sistema de justiça sobre os riscos e as limitações da IA também são essenciais. Juízes, advogados e policiais precisam estar cientes dos vieses potenciais e das implicações éticas associadas ao uso dessas tecnologias. Programas de formação podem ajudar a equipar esses profissionais com as habilidades necessárias para identificar e mitigar vieses, garantindo que a IA seja utilizada de maneira responsável e justa (Eubanks, 2018).

Outro aspecto relevante abordado por Bichara e Cascardo Junior (2023) é a necessidade de políticas públicas que proíbam o viés de confirmação em investigações policiais. Eles sugerem que a adoção de tais políticas pode melhorar significativamente a equidade e a justiça no sistema criminal, ao garantir que todas as evidências sejam consideradas de forma imparcial. Além disso, defendem que treinamentos contínuos para os investigadores sobre a identificação e a prevenção de vieses são fundamentais para a eficácia dessas políticas.

6. Considerações finais

A implementação da IA no sistema de justiça criminal oferece tanto oportunidades quanto desafios significativos. Embora a tecnologia tenha o potencial de melhorar a eficiência e a precisão das operações policiais e judiciais, é fundamental que se abordem os vieses e as injustiças inerentes ao uso desses sistemas. A prudência e a ética devem guiar a adoção de IA, com uma ênfase contínua na eliminação de preconceitos e na promoção da justiça equitativa.

Para evitar a perpetuação de um sistema de justiça racista e discriminatório, é essencial que as tecnologias de IA sejam desenvolvidas e implementadas com uma compreensão profunda das suas implicações sociais. A transparência, a responsabilidade e a inclusão de diversas perspectivas no processo de desenvolvimento são fundamentais para garantir que a IA contribua para um sistema de justiça mais justo e humanizado.

Em última análise, o uso da IA deve promover um ambiente onde a tecnologia complementa, mas não substitui a capacidade humana de julgar e decidir com base em princípios éticos e de equidade. Deve-se pensar não só no que a IA pode produzir de bom, mas também trabalhar para evitar aquilo que ela possa produzir de ruim. Somente por meio de um compromisso contínuo com a justiça social e a ética é que a IA poderá realizar seu verdadeiro potencial no sistema de justiça criminal.

Informações adicionais e declarações dos autores (integridade científica)

Declaração de conflito de interesses: os autores confirmam que não há conflitos de interesses na condução desta pesquisa e na redação deste artigo. **Declaração de autoria:** todos e somente os pesquisadores que cumprem os requisitos de autoria deste artigo são listados como autores; todos os coautores são totalmente responsáveis por este trabalho em sua

totalidade. **Declaração de originalidade:** os autores garantiram que o texto aqui publicado não foi publicado anteriormente em nenhum outro recurso e que futuras republicações somente ocorrerão com a indicação expressa da referência desta publicação original; eles também atestam que não há plágio de terceiros ou autoplágio.

Como citar (ABNT Brasil)

BICHARA, Anderson de Andrade; BRITO, Fabiana Amaro de. Desafios éticos ao uso da inteligência artificial no sistema de justiça criminal. **Boletim IBCCRIM**, São Paulo, v. 32, n. 383, p. 11-14, 2024. DOI: <https://doi.org/10.5281/zenodo.13630977>. Disponível em: https://publicacoes.ibccrim.org.br/index.php/boletim_1993/article/view/1242. Acesso em: 1 out. 2024.

zenodo.13630977. Disponível em: https://publicacoes.ibccrim.org.br/index.php/boletim_1993/article/view/1242. Acesso em: 1 out. 2024.

Referências

ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L. Machine bias: there's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, 23 maio 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 20 set. 2024.

ARISTÓTELES. *Ética a Nicômaco*. Tradução: Edson Bini. São Paulo: Edipro, 2009. Original publicado em 350 a.C.

BARABAS, Chelsea. Beyond bias: re-imagining the terms of "ethical AI" in criminal law. *Georgetown Journal of Law & Modern Critical Race Perspectives*, v. 12, n. 2, p. 83-111, 2020. <https://doi.org/10.2139/ssrn.3377921>

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. *Fairness and machine learning: limitations and opportunities*. Cambridge: MIT Press, 2023. Disponível em: <https://fairmlbook.org/pdf/fairmlbook.pdf>. Acesso em: 19 ago. 2024.

BICHARA, Anderson de Andrade; CASCARDO JUNIOR, Agostinho Gomes. Inquérito policial e proibição de viés de confirmação: rebatendo mitos histórico-doutrinários com base na Constituição e em Tratados Internacionais. *Jus*, 27 maio 2023. Disponível

em: <https://jus.com.br/artigos/104335/inquerito-policial-e-proibicao-de-vies-de-confirmacao-rebatendo-mitos-historico-doutrinarios-com-base-na-constituicao-e-em-tratados-internacionais>. Acesso em: 16 ago. 2024.

EUBANKS, Virginia. *Automating inequality: how high-tech tools profile, police, and punish the poor*. Nova York: St. Martin's Press, 2018.

KANT, Immanuel. *Groundwork for the metaphysics of morals*. Tradução: Allen W. Wood. New Haven: Yale University Press, 2002. Original publicado em 1785.

MILL, John Stuart. *Utilitarianism*. Oxford: Oxford University Press, 1998. Original publicado em 1863.

MITCHELL, Tom M. *Machine learning*. Nova York: McGraw-Hill Education, 1997.

NOBLE, Safiya Umoja. *Algorithms of oppression: how search engines reinforce racism*. Nova York: New York University Press, 2018.

RAWLS, John. *A theory of justice*. Cambridge: Harvard University Press, 1971.

Recebido em: 16.06.2024. Aprovado em: 30.07.2024. Última versão dos autores: 21.08.2024.