

Inteligência Artificial e o Desenvolvimento de Software: Vieses, Privacidade e Responsabilidade Ética

Breno Salu Souza de Oliveira¹, Elisa Ribeiro Gonçalves¹,
Emanuel Teixeira Martins¹, Rian Lucas da Silva Valadão¹,
Yasmin Souza Lima¹, Liziane Santos Soares¹

¹Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa Campus Rio Paranaíba
Rodovia BR 230 KM 7, Rio Paranaíba - MG - Brasil, 38810-000

{breno.salu, elisa.r.goncalves, emanuel.martins, rian.valadão,
yasmin.lima, liziane.soares}@ufv.br

Abstract. This paper addresses ethical issues in the development of Artificial Intelligence (AI) systems, with an emphasis on algorithmic bias, data privacy, and the social impact of these technologies. The technical, social, and legal implications of large-scale AI adoption are analyzed through case studies and recent research. Solutions such as promoting algorithmic transparency, strengthening the ethical training of professionals, and developing joint guidelines among academic institutions, government, and the private sector are discussed. It is emphasized that an interdisciplinary approach is essential to ensure that AI development is fair, responsible, and aligned with democratic values, contributing to a more ethical and inclusive technological future.

Resumo. Este trabalho aborda questões éticas na criação de sistemas de Inteligência Artificial (IA), com ênfase no viés algorítmico, privacidade dos dados e impacto social desses sistemas. As implicações técnicas, sociais e legais da adoção da IA em larga escala são analisadas através de cases e pesquisas recentes. São discutidas soluções como a promoção da transparência algorítmica, o fortalecimento da formação ética dos profissionais e a elaboração de diretrizes conjuntas entre instituições acadêmicas, governamentais e do setor privado. Destaca-se que uma abordagem interdisciplinar é essencial para garantir o desenvolvimento da IA de forma justa, responsável e alinhada aos valores democráticos, contribuindo para um futuro tecnológico mais ético e inclusivo.

1. Introdução

A Inteligência Artificial (IA) tem se consolidado como uma área central na transformação digital da sociedade, impactando áreas como saúde, justiça, segurança e mercado. O uso crescente de sistemas automatizados para tomada de decisão levanta questões éticas relevantes, especialmente quanto à transparência, equidade e privacidade. Modelos de IA, quando mal projetados, podem reforçar desigualdades e dificultar a responsabilização por decisões críticas.

No contexto brasileiro e internacional, a preocupação com o uso ético da IA tem gerado discussões regulatórias. [Gomes et al. 2021] destacam os limites éticos no uso

desses sistemas no Judiciário, chamando atenção para a necessidade de compatibilizar as inovações tecnológicas com marcos legais como a Lei Geral de Proteção de Dados (LGPD) [Garcia et al. 2020] e as resoluções do Conselho Nacional de Justiça (CNJ)¹ e a GDPR (*General Data Protection Regulation*)[European Union 2016]. A ausência de critérios claros para transparência algorítmica ameaça os direitos fundamentais em ambientes digitais.

Este artigo analisa os principais desafios éticos relacionados ao desenvolvimento de software com IA, com foco no viés algorítmico, na proteção de dados e na responsabilidade dos desenvolvedores. A proposta é articular essas questões ao contexto legal e às discussões contemporâneas sobre regulação e justiça tecnológica.

2. Desafios Éticos no Desenvolvimento de Sistemas com IA

A engenharia de software contemporânea, especialmente em domínios que empregam inteligência artificial (IA), transcendeu a mera automação de tarefas para se tornar uma força modeladora da sociedade. As decisões tomadas durante o ciclo de desenvolvimento de software, desde a concepção da arquitetura até a escolha de um modelo de aprendizado de máquina, carregam implicações éticas profundas que exigem uma análise rigorosa. O viés algorítmico não é uma falha espontânea dos sistemas, mas uma consequência direta de decisões humanas e dados históricos. As suas origens são primariamente duas: os dados de treinamento e as escolhas de modelagem. Dados de treinamento que refletem preconceitos e desigualdades sociais existentes ensinam o modelo a replicar e, por vezes, amplificar essas mesmas discriminações. Adicionalmente, as escolhas feitas por desenvolvedores, como a seleção de *features*, a definição da função objetivo ou a métrica de sucesso de um modelo, podem introduzir vieses de forma sutil.

Exemplos proeminentes e documentados na literatura ilustram a gravidade do problema. No campo do reconhecimento facial, o trabalho “*Gender Shades*” demonstrou que sistemas comerciais apresentavam taxas de erro drasticamente maiores para mulheres de pele escura em comparação com homens de pele clara, expondo um viés interseccional severo [Buolamwini and Gebru 2018]. Na justiça criminal, sistemas preditivos como o COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), utilizado para prever a reincidência criminal, foram alvo de críticas por exibirem um viés racial significativo, atribuindo falsamente taxas de risco mais altas a réus negros. Análises subsequentes confirmaram que, embora a precisão geral pudesse ser equivalente entre grupos raciais, o tipo de erro cometido era distribuído de forma desigual [Dressel and Farid 2018]. No setor de recrutamento, uma ferramenta de IA desenvolvida pela Amazon para análise de currículos foi descontinuada após ser identificado que o sistema penalizava candidatas mulheres, pois havia sido treinado com dados históricos predominantemente masculinos [Dastin 2018].

Além do problema de viés e discriminação, outro desafio ético central é a coleta de informação. A eficácia dos sistemas de IA modernos, especialmente os Grandes Modelos de Linguagem (LLMs), é diretamente proporcional ao volume e à diversidade dos dados disponíveis para seu treinamento. Essa enorme necessidade por dados impulsionou uma prática de varredura da internet em larga escala, onde empresas de tecnologia coletam volumes massivos de texto e código para que seus modelos “aprendam” o conhecimento

¹CNJ: <https://www.cnj.jus.br/>

humano. Esta abordagem, no entanto, opera em uma fronteira ética e legal complexa, desafiando as noções tradicionais de consentimento e propriedade intelectual.

Dois exemplos conhecidos ilustram este dilema. O primeiro é o *GitHub*, que hospeda milhões de repositórios de código-fonte públicos. Ferramentas de inteligência artificial, como o *GitHub Copilot*², foram treinadas utilizando esse vasto acervo de código, sem o consentimento explícito dos desenvolvedores para este fim específico. Isso gerou contestações significativas, pois muitos repositórios são regidos por licenças de código aberto (como a GPL) que impõem condições sobre a reutilização do código, condições que o treinamento de um modelo de IA comercial pode não satisfazer. A questão culminou em ações judiciais, como o processo *Doe v. GitHub*, que acusa as empresas de desrespeitarem as licenças e os direitos dos criadores, configurando uma forma de pirataria de software em escala industrial [Butterick and Joseph Saveri Law Firm & Craft 2022].

O segundo exemplo é o *Stack Overflow*³, uma plataforma comunitária baseada em perguntas e respostas que se tornou um pilar para desenvolvedores. Seu conteúdo, gerado por milhões de usuários, representa um *corpus* de altíssima qualidade para treinar IAs em lógica, programação e resolução de problemas. Recentemente, a plataforma formalizou parcerias para vender seus dados a grandes empresas de tecnologia para fins de treinamento de IA [Landymore 2024]. Essa decisão gerou um intenso debate na comunidade, pois os usuários que contribuíram com seu conhecimento para um bem comum não necessariamente consentiram com a comercialização de suas contribuições para o desenvolvimento de sistemas que podem, eventualmente, competir com a própria plataforma.

Essas práticas de coleta massiva de dados, seja de código ou de conversas técnicas, expõem a inadequação do modelo de consentimento baseado em “termos de serviço”. O dado, embora público, foi disponibilizado sob um contexto e propósito específicos, que são subvertidos quando utilizados para treinar um produto comercial proprietário. Embora marcos legais como o GDPR e a LGPD estabeleçam regras para o tratamento de dados pessoais, a aplicação dessas leis a dados “públicos” e a propriedade intelectual embutida neles, consiste em um grande desafio técnico e jurídico para os desenvolvedores [Christina Tikkinen-Piri 2017].

Em seu trabalho sobre vieses em sistemas computacionais, Friedman e Nissenbaum alertam que vieses podem ser difíceis de identificar e remediar. Eles argumentam que, especialmente em sistemas complexos, os vieses tendem a permanecer ocultos na estrutura do código, tornando sua identificação e explicação um grande desafio. Consequentemente, quando um sistema de computador discrimina algo de forma sistemática e injusta, a origem do problema pode ser obscurecida por decisões e limitações técnicas. Isso cria uma dificuldade em determinar a responsabilidade, pois, diferentemente de interações humanas, sistemas tendenciosos não fornecem mecanismos claros ou acessíveis para contestação ou apelação de suas decisões. [Friedman and Nissenbaum 1996].

Diante deste cenário de vieses, riscos à privacidade e dilemas de responsabilidade, a comunidade técnica e acadêmica tem proposto caminhos para mitigar esses problemas. Uma das frentes mais promissoras é a da Inteligência Artificial Explicável (XAI), cujo objetivo é desenvolver técnicas que tornem as decisões de modelos complexos, como

²GitHub e Copilot: <https://github.com/features/copilot>

³<https://stackoverflow.com/>

redes neurais profundas, mais transparentes e compreensíveis para os seres humanos [Adadi and Berrada 2018]. A explicabilidade é um pré-requisito para a auditoria e a depuração de vieses.

Em um nível mais amplo, organizações como o *Institute of Electrical and Electronics Engineers*, com sua iniciativa *Ethically Aligned Design* [IEEE 2017]; e a Organização para a Cooperação e Desenvolvimento Económico com seus princípios sobre IA [OECD 2019] publicaram guias e princípios para orientar o desenvolvimento responsável de tecnologias. Contudo, a efetividade desses guias depende de sua adoção pela indústria e, fundamentalmente, da formação dos próprios profissionais. Por isso, uma solução de longo prazo reside na inserção robusta da ética nos currículos de Ciência da Computação e Engenharia de Software, garantindo que a próxima geração de desenvolvedores esteja equipada não apenas com habilidades técnicas, mas também com a capacidade de raciocínio ético para construir sistemas mais justos e seguros.

3. Trabalhos relacionados

A ascensão da Inteligência Artificial representa uma das mais profundas transformações no desenvolvimento de software e na sociedade como um todo. À medida que algoritmos passam a desempenhar papéis centrais em decisões automatizadas, surgem questionamentos éticos que desafiam legislações existentes, princípios de justiça e a própria compreensão de responsabilidade tecnológica. A literatura especializada tem se dedicado a investigar essas questões sob múltiplas perspectivas, revelando o potencial da IA e os riscos associados ao seu uso indiscriminado ou mal regulado [Zag Sites 2025][Sobrinho 2025].

No contexto jurídico brasileiro, Gomes, Vaz e Dias [Gomes et al. 2021] discutem os desafios éticos impostos pela utilização de IA no sistema de justiça, especialmente diante da Lei Geral de Proteção de Dados (LGPD) e das resoluções emitidas pelo Conselho Nacional de Justiça. Os autores alertam para os perigos de se delegar decisões judiciais a sistemas automatizados sem uma supervisão humana adequada. Ressaltam, ainda, a importância de garantir a transparência dos algoritmos e de assegurar os direitos fundamentais dos cidadãos em um ambiente cada vez mais digitalizado.

Sob outra perspectiva, Buolamwini e Gebru [Buolamwini and Gebru 2018] analisam como sistemas de reconhecimento facial disponíveis no mercado apresentavam taxas de erro significativamente maiores ao identificar mulheres negras, em comparação a homens brancos. Essa disparidade revela como a ausência de diversidade nos conjuntos de dados e nas equipes de desenvolvimento pode levar a resultados discriminatórios. O estudo tornou-se um marco nos debates sobre justiça algorítmica, reforçando a urgência de incorporar princípios éticos desde as etapas iniciais da concepção de sistemas baseados em IA.

Selbst et al. [Selbst et al. 2019] avançam nesse debate ao apontar as limitações dos métodos tradicionais da Ciência da Computação frente a questões como justiça e equidade. Os autores discutem o chamado “erro de abstração”, que ocorre quando sistemas técnicos são projetados sem considerar adequadamente os contextos sociais nos quais serão aplicados. Tal desconexão compromete a eficácia de soluções que se pretendem justas, mas ignoram dinâmicas institucionais, históricas e culturais fundamentais.

Essa preocupação já havia sido sinalizada por Friedman e Nissenbaum (1996) [Friedman and Nissenbaum 1996], que classificaram os vieses computacionais em três

categorias: pré-existentes (de natureza social), técnicos (decorrentes do design) e emergentes (relacionados ao uso). Eles defendem que a ausência de imparcialidade nos sistemas compromete diretamente sua legitimidade social e eficácia ética. Dessa forma, a literatura converge ao destacar a urgência de repensar o desenvolvimento e a aplicação da IA a partir de uma abordagem ética, interdisciplinar e sensível à diversidade social. Princípios como explicabilidade, equidade e responsabilização devem estar no cerne de qualquer projeto que envolva inteligência artificial, especialmente quando seus impactos recaem sobre populações vulneráveis.

Os trabalhos analisados abordam aspectos relevantes da ética na IA, como vieses, transparência e limitações técnicas. No entanto, este trabalho se diferencia ao adotar uma perspectiva mais ampla relacionada a todos esses aspectos.

4. Conclusão

Os dilemas éticos que surgem com o uso cada vez maior da Inteligência Artificial no desenvolvimento de software são complexos e multifacetados, e demandam uma atenção rápida e cuidadosa. Questões como o viés nos algoritmos, a proteção de dados pessoais e a responsabilidade jurídica e social dos desenvolvedores deixaram de ser detalhes secundários e passaram a ser assuntos centrais na prática atual da engenharia de software.

Ao analisar essas questões, fica claro como decisões técnicas, muitas vezes consideradas neutras, podem ter impactos profundos na justiça, na igualdade e nos direitos fundamentais. Casos concretos mostram que sistemas mal planejados ou mal supervisionados podem aumentar desigualdades e colocar em risco valores democráticos. A dificuldade de entender como esses modelos funcionam, especialmente os baseados em aprendizado profundo, torna mais difícil identificar responsáveis pelos problemas e corrigir erros, reforçando a necessidade de métodos que sejam transparentes e passíveis de auditoria.

Sendo assim, é fundamental que a formação dos profissionais da área conte, de forma abrangente, a ética como elemento estruturante da prática computacional. Mais do que uma obrigação técnica, o desenvolvimento da IA deve ser visto como um pacto com a sociedade e com a política. Apenas através de uma perspectiva que envolva diversas áreas, que seja clara e que se baseie em princípios éticos será possível construir sistemas que realmente beneficiem a todos.

Referências

- [Adadi and Berrada 2018] Adadi, A. and Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, 6:52138–52160.
- [Buolamwini and Gebru 2018] Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency*, pages 77–91. PMLR.
- [Butterick and Joseph Saveri Law Firm & Craft 2022] Butterick, M. and Joseph Saveri Law Firm & Craft (2022). Github copilot litigation. Acesso em 19 jun. 2025. Disponível em: <https://githubcopilotlitigation.com/>.

- [Christina Tikkinen-Piri 2017] Christina Tikkinen-Piri, Anna Rohunen, J. M. (2017). Regulation (eu) 2016/679 of the european parliament and of the council. *Elsevier Science-Direct Journals*, 34(1):134.
- [Dastin 2018] Dastin, J. (2018). Amazon scraps secret ai recruiting tool that showed bias against women. Acesso em 19 jun. 2025. Disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scrapes-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G/>.
- [Dressel and Farid 2018] Dressel, J. and Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science advances*, 4(1):eaao5580.
- [European Union 2016] European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. General Data Protection Regulation (GDPR).
- [Friedman and Nissenbaum 1996] Friedman, B. and Nissenbaum, H. (1996). Bias in computer systems. In *Proceedings of the conference on Computers, freedom and privacy*, pages 21–30. ACM.
- [Garcia et al. 2020] Garcia, L. R., Aguilera-Fernandes, E., Gonçalves, R. A. M., and Pereira-Barreto, M. R. (2020). *Lei Geral de Proteção de Dados (LGPD): guia de implantação*. Editora Blucher.
- [Gomes et al. 2021] Gomes, E. B., Vaz, A. A., and de Oliveira Dias, S. M. (2021). Limites éticos para o uso da inteligência artificial no sistema de justiça brasileiro, de acordo com a lei 13.709 de 2018 (lgpd) e resoluções 331 e 332 do conselho nacional de justiça. *Revista Internacional CONSINTER de Direito*, 13:107–124. Acesso em 17 jun. 2025. Epub 08-Set-2022.
- [IEEE 2017] IEEE (2017). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 1*. IEEE.
- [Landymore 2024] Landymore, F. (2024). Stack overflow bans users for protesting against it selling their answers to openai as training data. The Byte. Acesso em 19 jun. 2025. Disponível em: <https://futurism.com/the-byte/stack-overflow-bans-users-openai-training-data>.
- [OECD 2019] OECD (2019). Oecd ai principles. <https://www.oecd.org/en/topics/ai-principles.html>. Accessed: 2025-07-03.
- [Selbst et al. 2019] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 59–68.
- [Sobrinho 2025] Sobrinho, G. (2025). Como a inteligência artificial está revolucionando o desenvolvimento de software em 2025. Acesso em 23 jun. 2025. Disponível em: https://sitedosobrinho.com/como-a-inteligencia-artificial-esta-revolucionando-o-desenvolvimento-de-software-em-2025/?utm_source=chatgpt.com.

[Zag Sites 2025] Zag Sites, K. A. S. M. (2025). O que é inteligência artificial? Acesso em 23 jun. 2025. Disponível em: https://zagsites.com.br/blog/o-que-e-inteligencia-artificial/?utm_source=chatgpt.com.