OS VIESES ALGORÍTMICOS E O DESAFIO ÉTICO DA INTELIGÊNCIA ARTIFICIAL

THE ALGORITHMIC BIASES AND THE ETHICAL CHALLENGE OF ARTIFICIAL INTELLIGENCE

Mariana Galdino Nunes Souza[1]

Resumo: Este resumo expandido explora os profundos vieses algorítmicos impostos à Inteligência Artificial (IA) na sociedade e seus impactos. Com base em uma abordagem descritivo-exploratória, a análise discute a inteligência e o aprendizado humano a partir de uma perspectiva pluralista, inspirada em obras literárias e científicas, para contextualizar a IA como uma ferramenta de inovação para o bem da sociedade. O estudo enfatiza a necessidade de abordagens éticas no desenvolvimento de softwares de IA, destacando a importância de uma abordagem multidimensional para mitigar vieses, desde a seleção de dados até auditorias e responsabilização adequada. Perspectivas diversas em conjuntos com dados e equipes de desenvolvimento são cruciais, assim como a implementação de treinamento contínuo e supervisão humana, visando transparência e a imparcialidade. Essas estratégias integradas promovem um desenvolvimento ético, transparente e equitativo da IA, essencial para o uso responsável em prol do bem-estar coletivo. O resumo expandido pôde concluir que a conscientização sobre vieses humanos e a adoção de práticas éticas são fundamentais para evitar discriminações e alinhar a tecnologia com elevados valores humanos.

Palavras-Chave: Inteligência Artificial; Vieses; Impacto Social; Ética.

Abstract: This expanded abstract explores the profound transformations imposed by Artificial Intelligence (AI) on society, focusing on algorithmic biases and their impacts. Based on a descriptive-exploratory approach, the analysis discusses human intelligence and learning from a pluralistic perspective, inspired by literary and scientific works, to contextualize AI as a tool for innovation for the good of society. The critical study emphasizes the need for ethical approaches in the development of AI software, highlighting the importance of a multidimensional approach to mitigate biases, from data selection to audits and adequate accountability. Diverse perspectives in datasets and development teams are crucial, as is the implementation of continuous training and human supervision, aiming for transparency and impartiality. These integrated strategies promote an ethical, transparent, and fair development of AI, essential for its responsible use for the collective well-being. The expanded abstract concluded that awareness of human biases and the adoption of ethical practices are fundamental to avoid discrimination and align technology with high human values.

Keywords: Artificial intelligence; Bias; Social Impact; Ethics.

INTRODUÇÃO

Admitindo as extensas transformações que a IA impõe na coletividade, como ao automatizar atividades cognitivas, analíticas e lógicas bem como acelerar o processamento de informações, este resumo molda-se como uma análise descritivo-exploratória com ênfase na discussão dos vieses algorítmicos e os impactos reais na coletividade.

[1] Graduanda em Direito pela FMU/SP.

Para essa finalidade, o estudo parte de um entendimento plurime de o que é a inteligência e o aprendizado, em especial o humano, base em análises de obras literárias e científicas, com o objetivo de projetar a IA como objeto de inovação para o bem comum. Nesse sentido, é enfatizada a necessidade de abordagens éticas no desenvolvimento e gestão de softwares de IA.

Ademais, destaca-se notoriedade de uma abordagem transversal para mitigar vieses, envolvendo, mas não se limitando a seleção e curadoria dos dados, auditorias recorrentes, responsabilização pelos atos cometidos, diversidade de perspectivas, treinamento contínuo bem como a supervisão humana. Desse modo, objetiva-se a promoção de um desenvolvimento ético, transparente e equitativo da IA para o bem-estar coletivo.

DISCUSSÕES E RESULTADOS

Numerosos pensadores estenderam suas metodologias para desenvolver trabalhos a fim de conhecer o processo de aprendizagem, como Sócrates que projetou a maiêutica (cada sujeito não teria conhecimento acrescido, onde cada resposta lhe fornecia nova pergunta); Piaget propôs estudos na observação de crianças, argumentou que o aprendizado por descobrimento, de forma autônoma, seria possível através de uma abordagem exposta a mais riscos, como o de tentar e errar. Já Paulo Freire propôs que o contexto em que o sujeito vive é essencial para concretizar ações pedagógicas, sendo essencial relação horizontal de interação entre educador e educando, através da troca de informações, “ninguém educa ninguém, como tampouco ninguém se educa a si mesmo: os homens se educam em comunhão, mediatizados pelo mundo” (FREIRE, 1983, p. 79).

Desse modo, a inteligência humana é compreendida como pluralista, envolvendo razão, emoção, biologia e sociedade, sendo o aprendizado um processo de adaptação e construção.

Em contraponto, a Inteligência Artificial opera de modo oposto, se acomodando próximo ao que Edgar Morin menciona como “inteligência cega”: uma máquina “"incapaz de contextualizar, destruindo os conjuntos e as totalidades, isolando todos os seus objetos do seu meio ambiente" (MORIN, 1999, p. 33, apud FERNANDES; GRAGLIA, 2024, p. 136). A IA fraciona o mundo dos dados em fatias "transforma o multidimensional em uma única dimensão" e, assim, "enfraquece a possibilidade de compreensão e reflexão" (MORIN, 1999, p. 33, apud FERNANDES; GRAGLIA, 2024, p. 136). A contradição ética está pronta: estamos substituindo o julgamento humano "complexo" (GARDNER) por uma "inteligência cega" (MORIN) e esperando que ela nos entregue um resultado justo.

É crucial entender, então, como essa "inteligência cega" opera tecnicamente. A IA é um ramo da computação focado em criar máquinas inteligentes, impulsionado por avanços tecnológicos. Ela funciona através de metodologias como o *Machine Learning*[2], diferente da práxis complexa (FREIRE) ou da inteligência contextual (GARDNER) , a máquina "aprende" por meio de algoritmos e dados, detalhados em tipos como o aprendizado, inclusive, não-supervisionado ou ainda por reforço.

O principal desafio ético surge com os vieses presentes nos algoritmos. Com base em Kahneman (sistemas 1 e 2 de pensamento), o autor explica que vieses cognitivos são distorções inconscientes que reforçam estereótipos. Consequentemente, esses vieses humanos são transmitidos aos algoritmos por meio de dados enviesados, resultando na amplificação de discriminações. Com finalidade ilustrativa, exemplos graves, como o sistema de recrutamento da Amazon que discriminou mulheres – o sistema foi treinado em dados a partir da base de colaboradores ativos, maioria masculina; (DASTIN, apud FERNANDES; GRAGLIA, 2024, p. 138). A IA aprendeu e entendeu que o sucesso se concentrava ao redor de homens e tomou os currículos femininos como inferiores, e por fim descartados.

O sistema judiciário não se isentou das falácias dos sistemas de autoaprendizado. Em 2016, o software COMPAS - Correctional Offender Management Profiling for Alternative Sanctions utilizado pelo poder judiciários nos EUA para identificar réus reincidentes foi alvo de investigação por parte do governo estadunidense.

---

[2] Vertente da IA que se dedica a mimetizar ações humanas, como a linguagem, através de padrões, por exemplo. Sem receber treinamento específico, executando de maneira automática para afiar sua precisão.

A apuração arguiu-se que o algoritmo atribuía “notas” desiguais a depender de diferentes grupos sociais, pois "dobravam as chances de atribuir uma nota negativa a uma pessoa negra em comparação a uma pessoa branca" (PROPUBLICA, 2016, apud FERNANDES; GRAGLIA, 2024, p. 139).

Importunamente é possível afirmar que a IA não estava errando, mas apenas executando de maneira impecável o status quo discriminatório que lhe foi induzido, aprendido com o histórico.

Um exemplo prático desses conflitos ocorre no Brasil. A IA aplicada em processos de concessão de benefícios pelo Instituto Nacional do Seguro Social (INSS) demonstrou o quanto decisões automatizadas podem ser questionadas, especialmente pela falta de clareza nos critérios adotados (SINSSP, 2023). Fica, assim, evidente que garantir uma supervisão humana contínua e um controle de qualidade robusto é essencial para prevenir

que as decisões sejam injustas ou discriminatórias.

A problemática não de desenvolve por trás da confiança irracional na tecnologia, mas em construir e rebuscar a governança de responsabilidade. Uma abordagem transversal para a mitigação, que é fundamentalmente um desafio ético.

Isso nos leva a estratégias reais e práticas. Iniciando pela governança e curadoria dos dados a serem imputados à IA, garantindo que sejam representativos e o livres de vieses. Em sequência, a formação de equipes multidisciplinares para garantir uma revisão equitativa e evitar vícios nos parâmetros. Além de promover o letramento da tecnologia, reforçando ideias éticas e livres de preconceitos históricos. Adicionalmente, as auditorias externas possuem um papel essencial ao verificar o que foi trabalho na linha de frente, dados inseridos nos sistemas e metodologias adotadas para treinamento da Inteligência Artificial.

Por fim, abordagens integradas fundamentais para promover transparência, equidade e responsabilidade.

CONCLUSÃO

Ao longo deste resumo expandido foi decorrido o risco representado pela IA em replicar preconceitos vivos na coletividade, considerando que seu aprendizado depende da fonte que a alimenta. A IA quando abandonada a própria sorte é aposta certeira a falha, amplificando É crucial avaliar as decisões algorítmicas para evitar que repliquem aversões. A resposta ao problema não se encerra com a curadoria dos dados, é preciso redesenhar o processo.

No caminho com destino à IA ética, é vital o "arremate" final das habilidades cognitivas humanas, não bastando a tecnologia por si. A responsabilidade de garantir que a tecnologia seja uma força benéfica é, e continuará sendo, completamente humana. Para que isso ocorra, é fundamental que estratégias integradas assegurem o desenvolvimento ético da IA. Isso inclui o uso de dados diversificados, a formação de equipes diversas, a implementação de supervisão humana, auditorias e treinamento contínuo.

Portanto, a conscientização sobre os vieses humanos, combinada com a ação coletiva, é essencial para garantir que a tecnologia seja uma força benéfica, alinhada com valores elevados e livre de discriminações.

REFERÊNCIAS

GARDNER, H. **Inteligências Múltiplas: a teoria na prática**. Trad. Maria Adriana Veríssimo Veronese. Porto Alegre: Artes Médicas, 1995.

MENESES, Isabel Rodrigues de. **Vamos limitar para criar?: uma análise sobre as exceções e as limitações do direito autoral brasileiro na era da inteligência artificial generativa**. Orientador: Anderson Souza da Silva Lanzillo. 2025. 129f. Trabalho de Conclusão de Curso - Curso de Direito, Centro de Ciências Sociais Aplicadas, Universidade Federal do Rio Grande do Norte, Natal, 2025.

FREIRE, P. **Pedagogia do oprimido**. 12. ed. Rio de Janeiro: Paz e Terra, 1983

KAHNEMAN, D. **Rápido e Devagar: duas formas de pensar**. Tradução de Cássio de Arantes Leite. Rio de Janeiro: Objetiva, 2012. [E-pub].

INSTITUTE FOR ELECTRICAL AND ELECTRONICS ENGINEERS. **Ethically Aligned Design: a Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems**. First Edition. IEEE, 2019.

SINSSP. IA no Setor Público (Parte 03): Conflitos e desafios da Inteligência Artificial no INSS. **Sinssp**, 24 out. 2023. Disponível em: https://sinssp.org.br/ia-no-setor-publico-parte-03-conflitos-e-desafios-da-inteligencia-artificial-no-inss/. Acesso em: 29 de outubro 2025.

MORIN, E. **O Pensar complexo: Edgar Morin e a crise da modernidade**. Org. Pena-Vega, Alfredo; do Nascimento, Elimar Pinheiro. 2ª ed. Rio de Janeiro: Garamond, 1999

JOHNSON, G.M. **Algorithmic bias: on the implicit biases of social technology**. Synthese 198, 9941–9961 (2021). Disponível em:

https://doi.org/10.1007/s11229-020-02696-y. Acesso em: 26 de outubro de 2025