

RESENHA CRÍTICA SOBRE OS VIESSES ALGORÍTMICOS PRESENTES EM SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

Diego Fernando Pereira Cavalcanti¹

Luiz Augusto Manfron Matias²

Rafael Luiz Cuypers³

Vinicius Rodrigues da Costa⁴

RESUMO

O artigo "Ethics and discrimination in artificial intelligence-enabled recruitment practices" de Z. Chen (2023) realiza uma análise aprofundada sobre a crescente implementação de sistemas de Inteligência Artificial (IA) em processos de recrutamento e seleção. O autor argumenta que, embora a IA prometa otimizar a contratação, aumentar a eficiência e reduzir tarefas repetitivas, ela carrega o risco intrínseco de perpetuar e amplificar vieses discriminatórios. A pesquisa adverte sobre os perigos do viés algorítmico, que pode institucionalizar práticas de contratação excludentes com base em gênero, raça, cor e personalidade. Chen identifica duas fontes principais para esse problema: o treinamento de algoritmos com conjuntos de dados históricos limitados e enviesados, que refletem preconceitos sociais preexistentes, e os próprios vieses dos desenvolvedores que projetam os sistemas. O artigo também explora a discriminação indireta, que ocorre quando o sistema utiliza variáveis correlacionadas (proxies), como CEP ou a universidade do candidato, para tomar decisões excludentes. Para enfrentar esses desafios, o autor propõe uma abordagem multifacetada, combinando soluções técnicas, como a criação de bancos de dados mais justos e o aumento da transparência algorítmica, com medidas gerenciais, como a implementação de governança ética corporativa e mecanismos de supervisão externa.

DISCUSSÃO CRÍTICA

A análise teórica de Chen é fortemente corroborada por exemplos concretos do mercado. O caso mais emblemático é o da ferramenta de recrutamento da Amazon, descontinuada após a empresa descobrir que o sistema penalizava sistematicamente currículos de candidatas mulheres. O algoritmo foi treinado com dados de contratação de um período de 10 anos, majoritariamente masculinos, o que o levou a associar sucesso profissional a candidatos homens. O sistema aprendeu a dar preferência a candidatos de universidades com histórico de ex-alunos do gênero masculino e a penalizar termos como "feminino" em currículos (e.g., "capitã do clube de xadrez feminino"). Este caso valida a tese central do "bias in, bias out" (viés que entra, viés que sai), demonstrando que o algoritmo apenas replicou com eficiência o preconceito existente nos dados de treinamento.

Outros exemplos notórios incluem a plataforma HireVue, que utiliza IA para analisar entrevistas em vídeo, avaliando expressões faciais, tom de voz e linguagem corporal. Pesquisas

1 Universidade Tuiuti do Paraná Curitiba – PR – Brasil. E-mail: diego.cavalcanti@utp.edu.br

2 Universidade Tuiuti do Paraná Curitiba – PR – Brasil. E-mail: luiz.matias@utp.edu.br

3 Universidade Tuiuti do Paraná Curitiba – PR – Brasil. E-mail: rafael.cuypers@utp.edu.br

4 Universidade Tuiuti do Paraná Curitiba – PR – Brasil. E-mail: vinicius.costa@utp.edu.br

revelaram que tais sistemas podem classificar com menor pontuação candidatos com sotaques, deficiências de fala ou pertencentes a minorias culturais, mesmo que possuam competências equivalentes. Da mesma forma, o sistema da Workday foi alvo de um processo judicial nos Estados Unidos sob a alegação de rejeitar automaticamente candidatos com base em idade e raça, utilizando variáveis correlacionadas para classificar os perfis sem transparência.

A crítica sobre vieses em sistemas de IA também é validada pela pesquisa "Gender Shades", de Joy Buolamwini e Timnit Gebru. O estudo demonstrou que sistemas de reconhecimento facial de grandes empresas apresentavam taxas de erro significativamente mais altas para mulheres de pele escura em comparação com homens de pele clara. Embora não seja exclusivamente uma ferramenta de recrutamento, essa tecnologia já é usada em fases de entrevistas por vídeo, e a pesquisa comprova cientificamente como tais sistemas podem falhar de forma discriminatória.

ANÁLISE DOS IMPACTOS SOCIAIS E ÉTICOS

Os impactos do uso de IA enviesada em processos seletivos são profundos. A consequência mais grave é a criação e o reforço de barreiras sistêmicas ao mercado de trabalho para grupos já marginalizados, cristalizando desigualdades sob um verniz de objetividade tecnológica. Ao automatizar e escalar preconceitos, as organizações não apenas perdem talentos valiosos, mas também obstruem a mobilidade social e a diversidade no ambiente corporativo.

Outro problema central é a opacidade de muitos desses sistemas, frequentemente descrita como o fenômeno da "caixa-preta" (black box). Muitas vezes, nem mesmo os desenvolvedores conseguem explicar certas decisões, o que impede a responsabilização e o direito à contestação. Ser rejeitado por um algoritmo sem uma justificativa clara é uma experiência desumanizadora, quemina a confiança dos candidatos nos processos de contratação e pode desmotivar profissionais qualificados.

Essa falta de transparência desafia marcos legais como o Regulamento Geral sobre a Proteção de Dados (GDPR) da União Europeia, que prevê um "direito à explicação" para decisões automatizadas. No entanto, a aplicabilidade prática desse direito ainda é tema de intenso debate, o que significa que a proteção contra decisões opacas não é garantida. Eticamente, a confiança cega em sistemas falhos sob o pretexto de eficiência representa uma grave abdicação da responsabilidade corporativa.

SUGESTÕES DE MELHORIAS

Para mitigar as injustiças, as soluções devem ser multifacetadas, combinando abordagens técnicas, organizacionais e regulatórias.

1. Melhorias Técnicas e de Dados: é crucial investir na construção de bancos de dados mais justos, representativos e balanceados, removendo proativamente informações que possam induzir a

correlações espúrias, como gênero e idade. A implementação de técnicas de Fairness-Aware Machine Learning (FAML), que incorporam restrições matemáticas para garantir equidade, é essencial. Além disso, deve-se aumentar a transparência algorítmica com ferramentas de explicabilidade (como LIME ou SHAP) que tornem as decisões comprehensíveis para recrutadores e candidatos.

2. Práticas Organizacionais e de Governança: A decisão final em processos críticos como a contratação deve permanecer sob supervisão humana (*human-in-the-loop*), com a IA funcionando como uma ferramenta de apoio, e não como o decisor final. As empresas devem implementar auditorias de impacto algorítmico regulares e independentes para identificar e corrigir vieses. Promover a diversidade nas equipes de desenvolvimento de tecnologia também é fundamental, pois grupos heterogêneos tendem a identificar vieses que poderiam passar despercebidos.

3. Regulamentação e Supervisão Externa: A criação de marcos regulatórios claros e indispensável. A Lei de Inteligência Artificial da União Europeia (EU AI Act), que classifica sistemas de recrutamento como de “alto risco”, é um exemplo proeminente, exigindo supervisão humana e transparência. Tais legislações precisam ser complementadas pela certificação de sistemas por órgãos independentes. Princípios de governança, como os propostos pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e adotados por países como o Brasil, também devem ser seguidos. Eles defendem que sistemas de IA devem respeitar os direitos humanos e valores como a equidade e a não discriminação.

CONCLUSÃO

A resenha do artigo de Chen, enriquecida por exemplos práticos e discussões éticas, confirma que a IA no recrutamento é uma faca de dois gumes. Seu potencial para otimizar processos é imenso, mas os riscos de aprofundar desigualdades são reais e já demonstrados. A automação acrítica ameaça perpetuar injustiças em uma escala sem precedentes.

O futuro da IA no mercado de trabalho não depende da rejeição da tecnologia, mas da nossa capacidade de desenvolvê-la e implementá-la de forma consciente, ética e responsável. A transição para um recrutamento mais tecnológico deve ser acompanhada por um compromisso firme com a transparência, a responsabilidade e a justiça. O desafio não é frear a inovação, mas guiá-la para que sirva à construção de um mercado de trabalho mais inclusivo e equitativo, em vez de reforçar as barreiras do passado. Em última análise, a inteligência dos sistemas que criamos será sempre um reflexo da nossa própria capacidade de agir com equidade e responsabilidade social.

REFERÊNCIAS

- BUOLAMWINI, J.; GEBRU, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Conference on Fairness, Accountability and Transparency, 2018, p. 77-91.
- CHEN, Z. Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities Social Sciences Communications*, v. 10, n. 567, 2023.
- DASTIN, J. Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*, 10 out. 2018.
- GOODMAN, B.; FLAXMAN, S. European Union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine*, v. 38, n. 3, p. 50-57, 2017.
- OECD. Recommendation of the Council on Artificial Intelligence (OECD AI Principles). OECD/LEGAL/0449, 2019.