

Presença de vieses algorítmicos em sistemas de Inteligência Artificial para recrutamento

Victor Hugo Piontkievitz da Cruz

¹Universidade Tuiuti do Paraná
Curitiba – PR

{victor.cruz}@utp.edu.br

Resumo. Com o grande progresso e conquistas na área de tecnologia, novas ferramentas surgiram, facilitando a realização de diversas atividades no dia a dia dos seres humanos e até mesmo trazendo a automação para diferentes tarefas. Uma dessas aplicações é vista no uso da Inteligência Artificial (IA) nos processos de seleção e recrutamento. No entanto, elas trazem consigo problemas relacionadas a discriminação e o preconceito, esses existentes por fatores como limitados conjuntos de dados brutos e dos projetistas de algoritmos serem tendenciosos, como diz a pesquisa de [Chen 2023], que busca entender e discutir as razões do viés algorítmico e forma que ocorre. A conclusão alcançada através de teoria fundamentada por meio de entrevistas e da revisão de literatura é que medidas técnicas como a transparência dos algoritmos, utilização de estruturas dados imparciais e medidas de gestões como uma governança ética da corporação e do auxílio de supervisão externa sejam aplicados.

1. Discussão crítica

A adoção de métodos e práticas de recrutamento envolvendo IA e Aprendizado de Máquina ou *Machine Learning* (ML) já está aplicado a cerca de 33% de organizações [Stephan 2017], permitindo que o trabalho seja melhorado em questões de facilidade, velocidade e recursos gastos. Além disso, muitas empresas de tecnologia veem criando e vendendo diversas ferramentas para permitir a identificação de candidatos sem viés, avaliar entrevistas de vídeos, potencializar a neurociência e a natureza dos jogos bem como para avaliar pessoas apropriadas para cargos de acordo com perfis em redes sociais [Gonzalez 2019]. Tais benefícios se demonstram desejados pelas diversas corporações, e são altamente utilizados hoje em dia pelos setores de Recursos Humanos (RH). A facilidade do trabalho permite a eficiência no processo de seleção, devido as capacidades de computação dos algoritmos utilizados, deixando que foco seja dado em outros aspectos dos responsáveis pelo setor, o que certamente pode ser compreendido.

A utilização de algoritmos é acreditada superar o trabalho humano quando se trata sobre o preconceito inconsciente por grandes empresas, como descoberto por [Beattie 2012]. Considerando a natureza humana, comportamentos preconceituosos podem ocorrer, sejam conscientes ou inconscientes, em diversas situações, incluindo durante a seleção de novos candidatos por parte do seletor. Como os algoritmos são considerados oferecem uma avaliação objetiva, e caso sejam objetivos, com a sua aplicação, esses elementos poderiam ser evitados. No entanto, os algoritmos são criados por humanos

e treinados sobre estruturas de dados, o que, dependendo do modo que são formados, lidados e analisados, podem resultar no viés algorítmico.

Segundo [Gonzalez 2019], os problemas obtidos do uso de IA e ML se referem a aplicações tradicionais, exemplificando o quanto é apropriado e a qualidade dos dados que estão sendo usados, bem como os resultados e decisões obtidos deles e os vieses presentes nas estruturas de dados que não respeitam parâmetros como a validação e justiça, indicando que os critérios dos algoritmos dependem da forma que são modelados e utilizados. [Gonzalez 2019] Também indica que os algoritmos observam as estruturas de dados, analisando as relações e outros padrões para definirem seus parâmetros de julgamento e classificação, para que possam ser realizadas previsões quando analisados os dados para as decisões. Considerando isso, é percebido que caso existam inconsistências nas estruturas de dados, tal vai ser a base para a avaliação do algoritmo, resultando em uma decisão enviesada de acordo com o que foi aprendido, afetando a objetividade e credibilidade do mesmo.

O problema do critério também é mencionado por [Gonzalez 2019], em que uma limitada consideração da mensuração da complexidade do critério pode resultar em uma baixa qualidade de julgamento e pouca captura dos elementos que realmente interessam, se comparando com as limitações das ferramentas de seleção tradicionais. Destaca-se então que o critério é um elemento fundamental que deve ser levado em consideração em relação aos algoritmos de recrutamento e seleção, bem como para diversos outros algoritmos. Os dados analisados devem ser focados em elementos significativos, permitindo que as decisões sejam realizadas corretamente e obtenham soluções positivas.

Como apontado por [Chen 2023], o viés pode existir por causa dos engenheiros do algoritmo, pois são eles que decidem detalhes desde os objetivos até a escolha dos modelos de dados utilizados para o ML, e caso essas pessoas possuam algum viés, este pode acabar sendo transmitidos para o algoritmo. Então, embora a avaliação das máquinas seja considerada objetiva, devido a forma em que são desenvolvidas e por serem altamente dependente das informações inseridas e implementadas nos algoritmos pelos desenvolvedores, detalhes como as características observadas, as relações entre os dados e os objetivos desejados podem acabar atrapalhando-a, trazendo resultados indesejados ou que possuam inconsistências, caso alguma forma de viés seja traduzido para a sua estrutura.

2. Análise dos impactos sociais e éticos

Os algoritmos são cada vez mais desenvolvidos e aperfeiçoados, sendo uma tecnologia que se tornou muito importante em várias áreas, incluindo o processo de seleção e recrutamento, no qual automatiza muitas das suas etapas. Primeiramente, as ferramentas desenvolvidas têm o objetivo de deixar os processos mais convenientes e eficientes em relações a diversos recursos. Comparados aos processos em que é realizada apenas a decisão humana, elas são consideradas a trazerem vantagens em questões de tempo, realizando uma rápida seleção entre grandes números de dados, em economia, em que é economizado recursos monetários e outros pela simplificação do processo, pela consideração de que as suas avaliações sejam mais precisas e observam importantes critérios, pela organização que podem oferecer ao usuário que as controla, além de que supostamente oferecem um julgamento objetivo.

No entanto, os algoritmos podem incluir um viés, obtido da forma que foram

modelados e desenvolvidos, podendo ser incorporado em sua estrutura em diversos aspectos. Isso pode resultar em um incorreto processo de seleção e de avaliação por parte do algoritmo, pois esse os realiza de acordo com o seu modelo e dos dados observados, que também podem conter inconsistências ou serem muito pequenos, insuficientes para realizar o julgamento correto, sendo refletidos na decisão final.

Com isso, os entrevistados pelo algoritmo no processo de seleção podem ser injustamente avaliados meramente por causa dessas em sua implementação, podendo receberem uma pontuação menor do que deveriam na percepção da máquina, correndo o risco de serem desconsiderados do processo ou subestimados, devido a indivíduos com maiores pontuações serem selecionados para as vagas ou o que quer que seja que o objetivo, enquanto isso possa passar despercebido pelo entrevistador que está utilizando as ferramentas de IA, pois este pode apenas estar operando o dispositivo em que o algoritmo é utilizado.

Como os algoritmos seguem suas estruturas de dados, caso o viés esteja presente em sua composição ou nos dados históricos, as chances de fazerem a seleção de forma incorreta pode continuar presente, inalterada, ou até mesmo piorada, caso não sejam corrigidos, criando uma tendência em relação ao modo em que avaliam cada indivíduo, prosseguindo para o futuro, gerando ainda mais dados enviesados, replicando o fenômeno de "viés para dentro e viés para fora" (Tradução do autor) [36KE. 2020].

Segundo [Chen 2023], o viés algorítmico se apresenta em termos de gênero, raça, cor e personalidade. Para gênero, em alguns casos é indicado que não são julgados neutralmente, devido a predominância do treinamento dos dados em pessoas do gênero masculino, resultando em discriminação contra pessoas do gênero feminino. Em relação a raça, entrevistados com nomes afro-americanos recebem cerca de 50% menos entrevistas que aqueles com nomes americanos. Em relação a cor da pele, cita o caso em que o dispensador de sabão estava falhando em reconhecer corretamente tons de peles, respondendo apenas a mãos brancas. A personalidade dos entrevistados também é julgada dependendo das palavras, tons de voz e expressões faciais, com emoções apresentadas sendo julgadas em positivas e negativas, e correlacionando curiosidade com a busca de novas oportunidades. Isso demonstra que devido ao viés algorítmico, essas características podem ser desnecessariamente e de forma incorretas levadas em consideração, afetando a avaliação final deles. Isso não só afeta as pessoas, como pode afetar a corporação que está utilizando dos algoritmos.

3. Sugestões de melhorias e regulamentações

Levando em consideração os vários motivos que podem levar a existência do viés algorítmico, existem algumas ações que podem ser adotadas para tentar mitigar o impacto causado. Para [Chen 2023], essas mudanças podem ser tanto em nível tanto quanto em nível regulamentário. Da perspectiva técnica, cita que a reconfiguração dos conjuntos de dados inconsistentes é uma das opções, utilizando múltiplos pontos de dados, porém pode ser uma ação custosa. Outro consiste em usar fontes de dados mais equitativas, permitindo uma justa avaliação, alterando e ajustando pontos de dados específicos durante o treino. Uma maior transparência dos algoritmos e o uso de ferramentas tecnológicas também é mencionado, pois uma maior visibilidade permite que as pessoas se tornem mais conscientes das possíveis discriminações, e muitos softwares que tentam deixar os

dados analisados mais imparciais, escondendo características ou identificando os problemas dos algoritmos.

Do ponto de vista de gestão, [Chen 2023] diz que as éticas sobre os algoritmos podem ser melhoradas. Os métodos mencionados vão da adoção de princípios voltados ao viés algorítmico, permitindo autorregulação, a utilização de auditorias sobre os sistemas algorítmicos para fazer ajustes caso preciso, incluindo diversos integrantes, permitindo a monitoração e correção do viés algorítmico, até o uso de supervisão externa, certificação e testes por terceiros para tentar mitigar os impactos negativos e a falta de confiabilidade. Esses métodos e princípios já são utilizados por diversas corporações, promovendo a equidade, ética e a responsabilidade.

Uma melhor monitoração e a adoção de princípios podem aprimorar a qualidade dos algoritmos em questão do viés, permitindo que os diversos elementos técnicos sejam ajustados e mais bem, além de que reforçá-los dentro das organizações promove um melhor senso de ética entre os membros. Em relação ao algoritmo, além da constante análise sobre sua objetividade, um melhor treinamento pode ser dado aos desenvolvedores, para que a importância seja dada para as corretas relações e critérios encontrados na sua estrutura.

4. Conclusão

A evolução das ferramentas tecnológicas se demonstra constante e ocorre de maneira rápida, com novas utilidades surgindo a cada dia em diversas áreas. A inteligência artificial possibilita que várias tarefas sejam realizadas mais rapidamente e eficientemente, além de outros benefícios, e sua aplicação em sistemas de recrutamento automatizados já é adotada por diversas organizações. No entanto, o viés algorítmico acaba surgindo na medida em que os algoritmos são criados de forma imparcialmente e dados inconsistentes são utilizados para o seu treinamento, esses quais geram ainda mais dados incoerentes. A discriminação se demonstra existente, causando-os a avaliarem os candidatos de maneira incorreta, podendo dar uma pontuação não correspondente as suas reais habilidades, o que pode resultar a esses entrevistados não serem selecionados para as vagas de emprego. Isso cria situações desfavoráveis não apenas para as pessoas candidatas, quanto para as empresas que as estão contratando. Para isso, medidas devem ser adotadas, como a implementação de regulamentações éticas nas empresas e um maior monitoramento sobre os algoritmos criados e dos dados utilizados para o treinamento das máquinas.

Referências

- 36KE. (2020). From sexism to recruitment injustice, how to make ai fair to treat.
- Beattie, G. & Johnson, P. (2012). Possible unconscious in recruitment and promotion and the need to promote equality. *Perspectives: Policy and Practice in Higher Education*, 16(1):7–13.
- Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities & Social Sciences Communications*.
- Gonzalez, M. F. e. a. (2019). "where's the i-o?" artificial intelligence and machine learning in talent management systems. *Personal Assesment and Decisions*, 5-3.
- Stephan, M. & Erickson, R. (2017). Talent acquisition: Enter the cognitive recruiter.