Resenha Crítica: Ética e Discriminação em Práticas de Recrutamento Habilitadas por Inteligência Artificial

Lucas Müller Scuzziato¹

¹Universidade Tuiuti do Paraná Curitiba – Paraná

lucas.scuzziato@utp.edu.br

Resumo. Esta resenha crítica aborda o artigo "Ethics and discrimination in artificial intelligence-enabled recruitment practices" de Zhisheng Chen, que investiga a discriminação algorítmica em processos de recrutamento impulsionados por Inteligência Artificial (IA) [Chen 2023]. O texto resume os principais pontos do artigo, discute criticamente exemplos reais de vieses algorítmicos em recrutamento e analisa os impactos sociais e éticos da IA em processos seletivos. Adicionalmente, são apresentadas sugestões de melhorias e regulamentações para mitigar injustiças e discriminações. Por fim, a resenha conclui com uma reflexão sobre o futuro da IA no mercado de trabalho, enfatizando a necessidade de uma abordagem ética e equitativa para o desenvolvimento e aplicação dessa tecnologia.

1. Introdução

A crescente integração da Inteligência Artificial (IA) em processos de Recursos Humanos (RH), particularmente no recrutamento, tem gerado debates significativos sobre seus benefícios e desafios éticos. O artigo "Ethics and discrimination in artificial intelligence-enabled recruitment practices" de Zhisheng Chen aborda essa lacuna de pesquisa sobre a discriminação algorítmica, explorando soluções técnicas e gerenciais para mitigar vieses em práticas de recrutamento impulsionadas por IA. Embora o recrutamento habilitado por IA tenha o potencial de aprimorar a qualidade do recrutamento, aumentar a eficiência e reduzir o trabalho transacional, ele também pode resultar em vieses algorítmicos que levam a práticas de contratação discriminatórias baseadas em gênero, raça, cor e traços de personalidade. Esta resenha crítica tem como objetivo aprofundar os pontos discutidos por Chen, apresentando um resumo do artigo, uma discussão crítica sobre exemplos reais de vieses, uma análise dos impactos sociais e éticos, e sugestões para melhorias e regulamentações, conectando o tema ao futuro da IA no mercado de trabalho.

2. Resumo do Artigo: Principais Pontos Discutidos

O estudo de Chen, fundamentado em uma revisão de literatura e na Grounded Theory, investiga o impacto da IA no recrutamento, focando nas causas e soluções para a discriminação algorítmica. A pesquisa aponta que o viés algorítmico decorre de conjuntos de dados brutos limitados e de designers de algoritmos tendenciosos. Para combater esses problemas, o artigo recomenda a implementação de medidas técnicas, como estruturas de conjuntos de dados imparciais e maior transparência algorítmica, bem como medidas gerenciais, incluindo governança ética corporativa interna e supervisão externa.

Os principais temas abordados no artigo são:

- 1. Benefícios do recrutamento baseado em IA: O recrutamento impulsionado por IA pode aprimorar a qualidade do recrutamento, aumentar a eficiência e reduzir o trabalho transacional. A IA pode coletar e avaliar grandes volumes de dados que excedem a capacidade analítica humana, permitindo recomendações de decisão. Além disso, sistemas de IA podem automatizar a triagem de currículos, resultando na redução de custos e carga de trabalho administrativa.
- 2. Causas da discriminação algorítmica: O viés surge principalmente de dados históricos incompletos ou tendenciosos, e da seleção de características pelos designers do algoritmo. Se os dados de treinamento são inerentemente injustos ou não representativos, o algoritmo replicará e amplificará esses preconceitos. A falta de diversidade na representação dos dados e as crenças pessoais dos engenheiros também contribuem para o viés.
- 3. **Tipos de discriminação algorítmica**: No processo de recrutamento, o viés algorítmico pode se manifestar em termos de gênero, raça, cor da pele e personalidade. Exemplos incluem a desvalorização de candidatas femininas e a associação de nomes de minorias étnicas a registros criminais.
- 4. Resolução da discriminação algorítmica: O estudo propõe soluções técnicas, como a construção de conjuntos de dados imparciais, melhoria da transparência algorítmica e uso de ferramentas tecnológicas contra vieses, como o *data blending* e testes diferenciais. Do ponto de vista gerencial, são sugeridas governança ética interna, auditorias regulares e supervisão externa, bem como a diversidade das equipes de desenvolvimento. Regulamentações e leis de proteção de dados também são apontadas como cruciais.

3. Discussão Crítica sobre Exemplos Reais de Vieses Algorítmicos em Recrutamento

A ascensão da inteligência artificial no recrutamento, embora promissora, trouxe à tona questões críticas sobre vieses algorítmicos. O artigo de Chen apresenta exemplos contundentes que ilustram como esses vieses se manifestam na prática, perpetuando e até amplificando desigualdades existentes.

Um dos casos mais notórios é o da Amazon, que em 2014 desenvolveu uma ferramenta de contratação baseada em IA. No entanto, o sistema exibiu um viés de gênero, desfavorecendo candidatas femininas. O problema surgiu porque o algoritmo foi treinado com currículos de funcionários predominantemente masculinos, refletindo um padrão histórico de contratação da empresa. Consequentemente, o algoritmo aprendeu que certas palavras ou expressões, mesmo que neutras, eram "masculinas" e desvalorizava currículos que continham termos associados a mulheres, como "feminina". Isso levou a Amazon a retirar a ferramenta e desenvolver um novo algoritmo imparcial.

Outro exemplo é o do chatbot Tay da Microsoft, que em 2016 aprendeu a produzir comentários sexistas e racistas no Twitter. Tay absorveu a linguagem dos usuários da plataforma como seus dados de treinamento. Em poucas horas, o chatbot inocente adotou discursos de ódio, forçando a Microsoft a desativá-lo. Este incidente demonstra que, quando as máquinas absorvem passivamente vieses humanos, elas podem refletir preconceitos inconscientes.

Além disso, a discriminação racial pode ser exacerbada por algoritmos de busca.

Pesquisas indicam que buscas por nomes associados a indivíduos negros eram mais propensas a serem acompanhadas por anúncios de registros de prisão, mesmo quando não existiam tais registros. Em contraste, buscas por nomes associados a indivíduos brancos não geravam tais anúncios. Um estudo sobre discriminação racial revelou que candidatos com nomes brancos recebiam 50% mais ofertas de entrevista do que aqueles com nomes afro-americanos.

No que diz respeito à cor da pele, em 2015, o algoritmo de um aplicativo de fotos do Google erroneamente rotulou uma foto de duas pessoas negras como gorilas. O algoritmo era insuficientemente treinado para reconhecer imagens com tons de pele escuros. O Google pediu desculpas e se comprometeu a prevenir tais erros. Da mesma forma, em 2017, um dispensador de sabão sem contato falhou em identificar corretamente tons de pele, respondendo apenas a mãos brancas e não detectando as mãos negras e pardas.

Esses exemplos sublinham a complexidade do problema. A crença de que a IA é inerentemente "objetiva" e "neutra" é um equívoco perigoso, que muitas vezes leva a que resultados discriminatórios sejam ignorados ou desconsiderados. Os vieses algorítmicos, sejam eles de gênero, raça, cor ou personalidade, refletem diretamente as desigualdades sociais e históricas presentes nos dados com os quais são treinados.

4. Análise dos Impactos Sociais e Éticos do Uso da IA em Processos Seletivos

O uso da IA em processos seletivos, apesar dos benefícios prometidos em termos de eficiência e redução de custos, levanta sérias preocupações sociais e éticas. A principal delas é a perpetuação e amplificação da discriminação. Algoritmos, ao serem treinados com dados históricos que contêm vieses implícitos ou explícitos da sociedade, podem replicar e escalar essas desigualdades de forma sistêmica e automatizada.

Um dos impactos sociais mais alarmantes é a **redução da equidade e oportuni- dades justas de emprego**. Quando um algoritmo é treinado com dados de uma força de trabalho predominantemente masculina, por exemplo, ele pode inadvertidamente desfavorecer candidatas mulheres, mesmo que qualificadas. Isso cria um ciclo vicioso onde grupos historicamente sub-representados são continuamente excluídos, minando os esforços de diversidade e inclusão nas empresas. A crença equivocada na "neutralidade" da IA
agrava o problema, pois as decisões algorítmicas tendem a ser vistas como imparciais, o que obscurece a discriminação e dificulta a responsabilização.

No âmbito ético, o uso de IA em recrutamento toca em questões fundamentais de **justiça social e direitos humanos**. A discriminação no mercado de trabalho é definida pela Convenção 111 da OIT como qualquer tratamento desfavorável baseado em raça, etnia, cor e gênero que prejudique a igualdade de emprego. Algoritmos que perpetuam esses vieses violam diretamente esse princípio. Além disso, a opacidade de muitos sistemas de IA (o chamado "problema da caixa preta") dificulta a compreensão de como as decisões são tomadas. Essa falta de transparência não apenas impede a responsabilização, mas também mina a confiança nos sistemas de IA.

Outro impacto ético é a **potencial invasão de privacidade e o uso indevido de dados**. Embora o artigo não se aprofunde em privacidade, a capacidade dos algoritmos de analisar grandes quantidades de dados para determinar "personalidade", tom de voz e

expressões faciais levanta questões sobre consentimento, segurança dos dados e o potencial para discriminação indireta baseada em informações não diretamente relacionadas às qualificações do trabalho.

A padronização excessiva e a falta de nuances individuais também são impactos sociais. Ao classificar candidatos com base em modelos de sucesso predefinidos, os algoritmos podem desconsiderar trajetórias de carreira não-convencionais, experiências diversas ou talentos emergentes que não se encaixam nos padrões históricos. Isso pode sufocar a inovação e limitar a diversidade de pensamento nas organizações.

Em suma, enquanto a IA promete eficiência e otimização, seu uso no recrutamento exige uma vigilância rigorosa para evitar que se torne um instrumento de injustiça e discriminação. Os impactos sociais e éticos da amplificação de vieses, da falta de transparência e da potencial invasão de privacidade são desafios que precisam ser enfrentados proativamente para garantir que a tecnologia sirva ao bem-estar humano e à equidade.

5. Sugestões de Melhorias e Regulamentações para Evitar Injustiças e Discriminações

Para combater os vieses algorítmicos e garantir um recrutamento justo e ético impulsionado pela IA, o artigo e a literatura correlata sugerem uma combinação de medidas técnicas e gerenciais, bem como a necessidade de regulamentações externas.

5.1. Medidas Técnicas

- Construção de Conjuntos de Dados Imparciais: A raiz do problema de viés algorítmico reside nos dados de treinamento tendenciosos. É crucial reconfigurar *datasets* desequilibrados , utilizando múltiplos pontos de dados e eliminando informações que reflitam preconceitos passados. Isso inclui a integração de "pequenos dados" com "grandes dados" para melhorar a precisão. Técnicas como o "oversampling" de dados incompletos ou o reamostragem de dados para reduzir vieses de categorização são essenciais.
- Aumento da Transparência Algorítmica: O "problema da caixa preta" impede a identificação de vieses. Aumentar a transparência significa que engenheiros e usuários devem ser capazes de entender como a IA chega a uma determinada decisão. Ferramentas como o *What-If tool* do Google e o *Fairness Flow* do Facebook auxiliam na detecção de vieses. A criação de "cartões de modelo" pode explicar o funcionamento, pontos fortes e fracos dos algoritmos.
- Ferramentas Tecnológicas contra Vieses: A utilização de softwares como o Blendoor, que "mescla" perfis de candidatos removendo nomes, fotos e datas para reduzir vieses inconscientes, é uma solução prática. A técnica de "decoupling" permite que o algoritmo identifique os melhores candidatos considerando variáveis otimizadas para grupos específicos, o que pode ajudar minorias. Além disso, a remoção de medidas específicas de gênero na incorporação de palavras pode reduzir o "viés de apresentação". Testes diferenciais, como o *Deep Xplore*, podem identificar vulnerabilidades em redes neurais.

5.2. Medidas Gerenciais

• Governança Ética Interna: Empresas de tecnologia devem estabelecer princípios de IA e comitês de padrões éticos para supervisionar e aplicar esses

- princípios. Auditorias internas regulares de IA-RH, semelhantes às auditorias tradicionais de RH, para revisar a seleção de funcionários e avaliar a confiabilidade dos algoritmos de IA e dados de ML são fundamentais.
- Supervisão Externa e Auditorias Regulares: A auto-regulamentação das empresas não é suficiente. É necessária a supervisão externa por meio de certificações de terceiros e testes de produtos de IA para garantir transparência e responsabilização. Auditorias regulares, envolvendo múltiplas partes na coleta de dados e monitoramento contínuo dos algoritmos, são cruciais para identificar e corrigir vieses.
- Diversidade na Equipe de Desenvolvimento: A composição demográfica da equipe que projeta os algoritmos é vital. A sub-representação de mulheres e minorias na indústria de tecnologia contribui para algoritmos tendenciosos, pois os algoritmos frequentemente refletem as opiniões de seus criadores. A inclusão de indivíduos diversos na coleta de dados e treinamento pode mitigar vieses enraizados.

5.3. Regulamentações e Leis

- Leis de Proteção de Dados e Não Discriminação: Leis como o GDPR da União Europeia são cruciais, pois abordam o impacto de algoritmos de Machine Learning e garantem um "direito à explicação" (Artigos 13-15), permitindo que os indivíduos solicitem explicações para decisões algorítmicas e exijam medidas para evitar influências discriminatórias em dados sensíveis. Além disso, o GDPR exige Avaliações de Impacto de Proteção de Dados (DPIAs).
- Iniciativas de Organizações Públicas: Entidades como a Algorithm Justice League (AJL) desempenham um papel importante ao delinear comportamentos vitais que as empresas devem seguir em acordos assináveis, responsabilizando aqueles que projetam e implantam algoritmos para melhorar as práticas.

6. Conclusão

O estudo de Chen demonstra que a Inteligência Artificial, embora promissora para otimizar o recrutamento, carrega o risco intrínseco de perpetuar e amplificar vieses e discriminações existentes no mercado de trabalho. A eficácia e a ética da IA dependem crucialmente da qualidade e imparcialidade dos dados de treinamento, bem como do cuidado com que é projetada e implementada.

O futuro da IA no mercado de trabalho é de crescimento contínuo, com a tecnologia prometendo transformar ainda mais os processos de identificação, avaliação e contratação de talentos. No entanto, para que a IA atinja seu potencial pleno e beneficie a todos, é imperativo que o foco se desloque da mera eficiência para a equidade e a responsabilidade. Se os vieses algorítmicos não forem mitigados, a IA, em vez de ser uma ferramenta de progresso, poderá aprofundar as desigualdades sociais e econômicas.

A necessidade de construir conjuntos de dados imparciais , aumentar a transparência dos algoritmos e implementar governança ética rigorosa são passos cruciais para construir sistemas de IA justos e confiáveis. Além disso, a diversidade nas equipes de desenvolvimento de IA e a implementação de regulamentações eficazes são fundamentais para assegurar que a tecnologia sirva aos valores humanos de justiça, igualdade e

inclusão. A capacidade de transcender os vieses históricos e criar um mercado de trabalho mais equitativo será o verdadeiro teste do potencial transformador da Inteligência Artificial. Se abordarmos esses desafios de frente, a IA poderá, de fato, ser um motor para um futuro mais justo e eficiente para o trabalho.

Referências

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities Social Sciences Communications*, 10(1):567.