Revista ARACÊ
ISSN: 2358-2472

PRECONCEITO DE GÊNERO NA LINGUÍSTICA CRÍTICA: DISCRIMINAÇÃO ALGORÍTMICA CONTRA PESSOAS TRANSGÊNERAS EM SISTEMAS DE POLÍTICAS PÚBLICAS

GENDER BIAS IN CRITICAL LINGUISTICS: ALGORITHMIC DISCRIMINATION AGAINST TRANSGENDER PEOPLE IN PUBLIC POLICY SYSTEMS

PREJUICIO DE GÉNERO EN LA LINGÜÍSTICA CRÍTICA: DISCRIMINACIÓN ALGORÍTMICA CONTRA LAS PERSONAS TRANSGÉNERAS EN LOS SISTEMAS DE POLÍTICAS PÚBLICAS

https://doi.org/10.56238/arev7n8-047

Data de submissão: 07/07/2025 **Data de publicação:** 07/08/2025

Claudio Noel de Toni Junior
Pós-Doutor em Geografia
Instituição: Universidade Federal de São Carlos (UFSCar)
E-mail: juniortoni100@gmail.com

RESUMO

O artigo investiga o preconceito de gênero na linguística crítica, focando como os algoritmos de inteligência artificial (IA) reproduzem e ampliam estereótipos contra pessoas transgêneras em áreas prioritárias como saúde, educação e emprego. Ao analisar, interdisciplinarmente, as relações entre linguagem, poder e tecnologia, o principal objetivo do estudo é identificar os mecanismos que levam à discriminação algorítmica e discutir abordagens críticas para mitigar tais vieses. A metodologia enfatiza métodos de *debiasing*, sendo uma técnica para corrigir preconceitos em modelos informacionais que usam a IA.

Palavras-chave: Linguística Crítica. Preconceito de Gênero. Inteligência Artificial (IA). Discriminação Algorítmica.

ABSTRACT

The article investigates gender bias in critical linguistics, focusing on how artificial intelligence (AI) algorithms reproduce and amplify stereotypes against transgender people in priority areas such as health, education, and employment. By analyzing, in an interdisciplinary manner, the relationships between language, power, and technology, the main objective of the study is to identify the mechanisms that lead to algorithmic discrimination and discuss critical approaches to mitigate such biases. The methodology emphasizes debiasing methods, which are techniques for correcting biases in informational models that use AI.

Keywords: Critical Linguistics. Gender Bias. Artificial Intelligence (AI). Algorithmic Discrimination.

RESUMEN

El artículo investiga los prejuicios de género en la lingüística crítica, centrándose en cómo los algoritmos de inteligencia artificial (IA) reproducen y amplían los estereotipos contra las personas transgénero en áreas prioritarias como la salud, la educación y el empleo. Al analizar, de manera interdisciplinaria, las relaciones entre el lenguaje, el poder y la tecnología, el objetivo principal del estudio es identificar los mecanismos que conducen a la discriminación algorítmica y discutir enfoques

Revista ARACÊ
ISSN: 2358-2472

críticos para mitigar tales sesgos. La metodología enfatiza los métodos de debiasing, una técnica para corregir los prejuicios en los modelos informativos que utilizan la IA.

Palabras clave: Lingüística Crítica. Prejuicio de Género. Inteligencia Artificial (IA). Discriminación Algorítmica.

Revista ARACÊ
ISSN: 2358-2472

1 INTRODUÇÃO

A interseção entre Linguística, Inteligência Artificial (IA) e as práticas de inclusão social tem emergido como um campo profícuo para a promoção de abordagens equitativas em setores críticos, como na saúde, na questão do emprego, além da Educação. Este artigo pretende examinar a importância dos estudos de Linguística na mitigação de preconceitos presentes em algoritmos de IA, com foco na comunidade trans, e articular como os conceitos teóricos podem contribuir para a compreensão e transformação das práticas discriminatórias embutidas nos sistemas tecnológicos.

Estudos na área de linguística crítica têm demonstrado, há décadas, que a linguagem não somente reflete, mas também reproduz e reforça relações de poder nas sociedades (Taso *et al*., 2023). Atualmente, no contexto digital, essa análise adquire novas dimensões, pois os algoritmos de IA, onde conjuntos de regras e processos computacionais que permitem que máquinas realizem tarefas de maneira inteligente, têm um papel decisivo na mediação de interações e representações sociais. Essa articulação, em particular, se torna problemática para grupos cujas identidades de gênero não se encaixam nos padrões tradicionais, como acontece com pessoas transgêneras e identidades não-binárias.

A partir de uma revisão da literatura que aborda os temas do preconceito algoritmo nas relações institucionais de saúde e emprego, principalmente, propõe-se investigar como as IA reproduzem estereótipos e excluem identidades diversas. Entre as principais questões que orientam este trabalho, destaca-se: como os algoritmos, ao processar a linguagem natural, incorporam preconceitos de gênero que já existem em *corpora* (coleções de textos) históricos e sociais? Qual é o impacto dessa discriminação algorítmica na vida de pessoas transgêneras e não binárias? E, finalmente, quais estratégias metodológicas e tecnológicas podem ser adotadas para mitigar esses vieses?

O objetivo principal desta investigação é desvelar e compreender os aspectos intrínsecos ao preconceito de gênero na linguística crítica, enquanto se evidencia a exclusão de identidades trans e não binárias em sistemas algorítmicos. Assim, se almeja oferecer subsídios teóricos e metodológicos que possam orientar pesquisadores da área, bem como desenvolvedores e a sociedade em geral na construção de tecnologias mais inclusivas.

O artigo delimita o *corpus* analisado aos algoritmos de processamento de linguagem natural (PLN) que se refere ao uso de inteligência artificial para entender e gerar linguagem humana e aos geradores de imagem que, historicamente, têm gerado produções discriminatórias (Salmoria e Ramos, 2024; Dev *et al*., 2021).

Revista ARACÊ
ISSN: 2358-2472

2 METODOLOGIA

A abordagem metodológica deste estudo combinou métodos qualitativos e quantitativos. Inicialmente, foi realizada uma análise documental dos principais estudos que investigam a interseção entre a linguística crítica e a IA, com ênfase nos mecanismos de exclusão presentes nos algoritmos. Em seguida, foi delimitado um *corpus* linguístico composto por textos gerados por sistemas de processamento de linguagem natural e algoritmos de imagem, conforme identificado na literatura (Salmoria e Ramos, 2024; Taso *et al.,* 2023).

Mediante pesquisa em base de dados acadêmico, como o *Google Scholar*, *Scielo* e base de dados de universidades internacionais, com destaque para o protagonismo dos estudos nesta área de pesquisa, percebe-se que a maioria da literatura provém de países da Europa ocidental e dos Estados Unidos.

Com base em estudos anteriores (Gonen e Goldberg, 2019), a segunda etapa consistiu em aplicar métodos de *debiasing*[1] (que visam reduzir preconceitos presentes em dados) em *embeddings*[2] de palavras (representações matemáticas de palavras), visando avaliar a eficácia desses procedimentos para mitigar os vieses identificados e por fim a relação humana com a máquina, imprencíndivel como forma sinérgica de alternativa de correção de dados.

3 CONTEXTO HISTÓRICO NA PERPECTIVA DA LINGUÍSTICA CRÍTICA

A linguística crítica, como campo de estudo, tem uma de suas missões centrada na análise das dimensões ideológicas da linguagem. Desde seus primórdios, essa abordagem utiliza metodologias que revelam como as estruturas discursivas contribuem para a manutenção de desigualdades sociais. Ao focar no preconceito de gênero, pesquisadores demonstram que os estereótipos não emergem de maneira neutra; ao contrário, são construídos e reproduzidos historicamente em múltiplos níveis (Taso *et al*., 2023).

Quando esses estereótipos se infiltram nos algoritmos de IA, a consequência é a perpetuação de uma lógica excludente que penaliza aqueles que não se encaixam nos parâmetros binários.

Estudos recentes indicam que os algoritmos de IA não são meros instrumentos neutros. Pelo contrário, eles operam com base em dados que refletem valores e vieses historicamente estabelecidos. Por exemplo, Salmoria e Ramos (2024) demonstraram como modelos de IA em geradores de imagem

[1] Ao traduzir para o português, o significado é desvirtualizado, visa fazer com que os dados informacionais disseminados sejam cada vez mais coesos e com menor nível de desigualdade de gênero nas relações interpessoais que usam sistemas de informação como a IA.
[2] *Embeddings* que tem o significado de "incorporar" é usado em práticas de Processamento de Linguagem Natural (PLN), sendo um vetor para embutir informações complexas, como documentos, palavras ou frases em algoritmos.

Revista ARACÊ
ISSN: 2358-2472

tendem a reproduzir estereótipos que marginalizam identidades trans e *queer*, evidenciando um quadro de exclusão que se reflete em práticas sociais discriminatórias. Essa constatação reforça a necessidade de implementar práticas rigorosas de *debiasing* e de reformular os conjuntos de dados utilizados nos treinamentos de algoritmos.

Do mesmo modo, De Lima Viana (2023) salientou que os algoritmos utilizados em processos seletivos e em contextos de saúde reproduzem desigualdades, impactando negativamente as pessoas trans. Esses estudos destacam que a inteligência artificial, ao filtrar e interpretar a linguagem, reforça uma estrutura de poder na qual apenas determinadas maneiras de manifestação de gênero são consideradas legítimas.

Nesta perpectiva, a análise da linguagem revela como discursos de poder são codificados e reiterados em diversas esferas, inclusive na tecnologia. Como destacado por Taso *et al*. (2023), os algoritmos que operam sobre corpora agrupam traços das estruturas sociais que os originaram. A exclusão de identidades não binárias e trans nos sistemas de classificação e priorização pode ser entendida como a materialização de um preconceito que historicamente privilegia formas linguísticas e visuais associadas a uma norma binária.

Nesse sentido, a aplicação dos métodos de estudos da língua crítica vai além da simples identificação de palavras ou imagens problemáticas; ela envolve também a análise do contexto histórico e sociocultural do ambiente social e cultural que influencia a linguagem em que esses elementos estão inseridos. A análise crítica torna-se, assim, um instrumento essencial para compreender não somente o que é discriminado, mas também por que e como ocorre esse processo de exclusão.

Para Kotec; Dockun e Sun (2024), ao adotar esta perspectiva, possibilita uma leitura aprofundada dos dados e oferece suporte para a reformulação dos algoritmos. Pela análise dos discursos embutidos nos sistemas de IA, é possível identificar pontos de inflexão e as mudanças significativas nos processos algorítmicos, proporcionando caminhos práticos para a intervenção e a promoção de uma maior equidade.

Embora o conjunto de textos analisados selecionado para este estudo tenha sido delimitado para abranger somente tecnologias que demonstraram evidências de preconceito de gênero em sistemas de políticas públicas: qual sejam: saúde, educação e emprego, reconhece-se que essa delimitação não é isenta de desafios. Entre os principais limites estão a heterogeneidade dos dados representados pela diversidade das informações e a dificuldade em determinar a generalização, ou seja, a aplicação dos resultados e dos achados para outros contextos tecnológicos.

Revista ARACÊ
ISSN: 2358-2472

Por um lado, a concentração em algoritmos de processamento de linguagem natural e geradores de imagem permite uma análise detalhada dos casos mais críticos (Salmoria e Ramos, 2024). Essa abordagem restringe a abrangência da investigação, uma vez que outras áreas da IA, como sistemas de recomendação e análise de sentimentos, também podem reproduzir vieses na discriminação do gênero.

Conforme Seco (2024) e Gonen e Goldberg (2019), o desafio de mensurar o viés quantitativamente impõe restrições metodológicas significativas, os métodos de *debiasing* aplicados em *embeddings* de palavras, embora úteis, apresentam limitações que podem mascarar, mas não eliminar os preconceitos subjacentes. Dessa forma, a interpretação dos dados exige uma atenção contínua à complexidade dos mecanismos envolvidos na discriminação algorítmica.

O estudo da exclusão de gênero em algoritmos e sistemas de IA no Brasil demonstra que o viés de gênero é um desafio multifacetado, que perpassa diversas esferas, desde o recrutamento de talentos até a administração dos serviços públicos e evidência, como a implementação de técnicas de *debiasing*, por si só, não resolve a problemática, exigindo intervenções integradas que envolvem auditorias regulares, a promoção de equipes diversificadas e uma governança transparente dos algoritmos.

Em um contexto em que os dados e algoritmos moldam as decisões de instituições públicas e privadas, a responsabilidade de mitigar o preconceito de gênero não pode repousar somente sobre os ombros dos desenvolvedores de IA.

No cenário contemporâneo, os algoritmos de inteligência artificial são amplamente utilizados em áreas que impactam diretamente a vida das pessoas. Contudo, a incorporação de preconceitos sociais e vieses linguísticos nesses sistemas pode agravar desigualdades já existentes e afetar negativamente comunidades historicamente marginalizadas. Estudos recentes apontam que os modelos de linguagem podem perpetuar estereótipos de gênero e outros estigmas, ressaltando a necessidade de uma abordagem crítica que envolva diferentes campos do conhecimento (Araújo e Araújo, 2024; Caliskan, Bryson e Narayanan, 2017).

A discriminação de gênero em sistemas de IA não é um fenômeno recente. Relatos de viés em algoritmos utilizados tanto no recrutamento quanto em contextos de saúde, educação e emprego evidenciam a necessidade de monitoramento constante das tecnologias emergentes. Estudos apontam que algoritmos podem reproduzir e até acentuar preconceitos historicamente presentes na sociedade (Fernandes e Graglia, 2024). Considerando os diferentes contextos de aplicação, é importante entender que o preconceito de gênero se manifesta em múltiplas dimensões, muitas vezes sutilmente, mas com consequências profundas para grupos minoritários.

Em sistemas de recrutamento, por exemplo, o viés pode ocorrer devido à predominância de dados históricos que refletem práticas discriminatórias. Fernandes *et al.* (2022) demonstram que a

Revista ARACÊ
ISSN: 2358-2472

utilização de soluções de aprendizagem de máquina, processando currículos com características masculino-cêntricas, reproduziu a exclusão de mulheres em processos seletivos, apesar da existência de métodos de *debiasing*. Eles ressaltam ainda que a intervenção humana é imprescindível para interpretar resultados e ajustar as práticas de seleção, garantindo a inclusão de mulheres no mercado de trabalho.

Em um cenário socioeconômico complexo como o brasileiro, a aplicação desses algoritmos em instituições tanto públicas quanto privadas pode acentuar desigualdades sociais. A persistência dos preconceitos mesmo diante de medidas de *debiasing* levanta questionamentos sobre a eficácia desses métodos e sobre a necessidade de práticas que transcendem a mera aplicação de técnicas computacionais de correção.

Para Fernandes *et al*., (2022), a adoção de algoritmos na análise de currículos promete rapidez na triagem dos candidatos, contudo, quando os dados históricos são enviesados, o resultado tende a vincular os mesmos preconceitos que ocorreram, por exemplo, em instituições de ensino e empresas ao longo do tempo.

Um estudo de caso discutido na literatura internacional, mas que remete às preocupações globais e encontra ecos no contexto brasileiro, é o algoritmo utilizado pela Amazon. Este algoritmo foi treinado com base em currículos dominados por homens e, consequentemente, desenvolveu um viés contra candidatas mulheres, penalizando currículos que explicitassem a experiência ou a formação em ambientes predominantemente femininos. Embora o exemplo da Amazon não seja oriundo do Brasil, ele ilustra claramente como dados enviesados podem conduzir à exclusão de grupos minoritários, uma realidade potencialmente reproduzida em outros contextos, inclusive em bancos de dados de instituições brasileiras.

A análise dos sistemas de recrutamento e seleção em instituições brasileiras revela que o preconceito de gênero não é uma anomalia pontual, mas sim um reflexo das estruturas sociais historicamente desiguais. Fernandes *et al*. (2022) enfatizam que, mesmo com a automação dos processos de seleção, os dados utilizados frequentemente reproduzem estereótipos e talheres discriminatórios que excluem candidatas qualificadas. Um exemplo narrado na literatura internacional, o algoritmo da Amazon torna-se ilustrativo nesse contexto. Quando um modelo é treinado com dados majoritariamente representativos dos currículos masculinos, ele passa a rebaixar a relevância daqueles que carregam marcadores associados ao feminino, como experiências advindas de instituições exclusivas para mulheres ou o simples uso da palavra “mulher” no currículo.

Mesmo quando técnicas de *debiasing* são implementadas, estudos indicam que estes procedimentos podem reduzir, mas não eliminar integralmente os vieses existentes. A intervenção

Revista ARACÊ
ISSN: 2358-2472

humana no processo de seleção é, portanto, um componente necessário, ao permitir que os resultados sejam interpretados de forma crítica e ajustados segundo os princípios da não discriminação e da transparência, conforme proposto pela Lei Geral de Proteção de Dados (LGPD) conforme (Santos e Graminho, 2024).

Instituições que administram bancos de dados com informações de pacientes e tratamentos podem recorrer a algoritmos para auxiliar na tomada de decisão clínica, recomendação de tratamentos ou até mesmo na priorização de recursos. Nesses contextos, a exclusão de gênero pode se materializar na forma de diagnósticos imprecisos, tratamentos inadequados ou mesmo na marginalização de grupos de pacientes.

Caso um sistema de IA for treinado com dados que contêm sub-representações de mulheres trans, principalmente, em determinadas condições clínicas, há o risco de que o algoritmo não identifique corretamente sintomas ou recomendações efetivas para o público feminino.

Esse cenário é especialmente preocupante em contextos de saúde, onde a precisão e a equidade no atendimento são cruciais para a eficácia dos serviços prestados. A aplicação de técnicas de *debiasing* pode ajudar a identificar e corrigir esses vieses, mas caso não sejam acompanhadas de uma auditoria contínua e de um acompanhamento humano, o risco de perpetuação da discriminação permanece (Fernandes e Graglia, 2024).

Rafaela Silva (2024) aborda que a falta de diversidade nas equipes que desenvolvem essas tecnologias pode resultar em uma visão limitada sobre as necessidades e as peculiaridades dos diferentes gêneros. Assim, mesmo em sistemas que buscam aplicar métodos de correção de vieses, a ausência de uma abordagem multidisciplinar pode comprometer a eficácia das soluções implementadas.

Nos sistemas de saúde, a exclusão de gênero pode ter implicações ainda mais graves, uma vez que a má interpretação de dados pode levar a diagnósticos errôneos, tratamentos inadequados e, consequentemente, a impactos negativos na qualidade de vida dos pacientes. Em contextos onde os algoritmos são empregados para estabelecer prioridades no atendimento médico ou para sugerir intervenções terapêuticas, a reprodução de dados enviesados pode comprometer a efetividade das políticas de saúde.

Instituições que operam com vastos bancos de dados clínicos podem estar sujeitas à sub-representação de mulheres trans em determinadas categorias de diagnóstico ou na avaliação de riscos. Ao treinar sistemas com dados que não refletem fielmente a diversidade dos pacientes, os algoritmos podem apresentar resultados que desconsideram as especificidades de gênero nas manifestações clínicas. Pode-se mencionar, doenças com sintomas que variam conforme o gênero, podem ser

Revista ARACÊ
ISSN: 2358-2472

interpretadas incompletamente se baseadas em dados predominantemente masculinos, comprometendo assim a correta identificação e tratamento para mulheres trans, que nasceram em um corpo masculino, onde a estrutura hormonal é diferenciada das mulheres cis, inclusive em relação ao uso do nome social que não é identificado pelo algoritmo.

Mesmo quando são implementadas estratégias de d*ebiasing*, a eficácia desses métodos depende da qualidade e da representatividade dos dados originais. É recomendado a aplicação de técnicas de correção nas instituições públicas e privadas e que possam adotar políticas de coleta e atualização constante dos bancos de dados, garantindo que as informações utilizadas reflitam as realidades diversas do público e privado atendidos. Técnicas de *compliance* e *accountability* e a responsabilidade no uso dos algoritmos, conforme preconizado pela LGPD, são fatores determinantes para mitigar o risco de exclusão e assegurar um atendimento equitativo (Santos e Graminho, 2024).

A utilização de técnicas de *debiasing* tem sido uma estratégia amplamente adotada para combater discriminações de gênero em algoritmos. Tais técnicas consistem em ajustes de modelos e conjuntos de dados para minimizar vieses e promover a equidade (Saleiro *et al*., 2018). No entanto, a literatura aponta que, embora o *debiasing* possa reduzir os efeitos discriminatórios, ele tem raramente o poder de eliminar as desigualdades embutidas nos dados.

Em processos seletivos, mesmo após a implementação de métodos de correção, a presença de viés residual decorrente de dados históricos tendenciosos ou da complexidade dos fatores socioculturais envolvidos permaneceu evidente nos resultados dos algoritmos (Fernandes *et al.*, 2022).

A linguística aplicada aos estudos dos algoritmos de IA é uma ferramenta que deve identificar os estereótipos e preconceitos presentes nos modelos de linguagem. Pesquisas indicam que os algoritmos podem reproduzir associações de gênero e outras classificações enviesadas, que refletem esferas históricas de poder e dominação (Caliskan; Bryson, e Narayanan, 2017). Tal perspectiva reafirma a necessidade de uma intervenção crítica que busque a revisão e implementação de dados representativos e linguagens inclusivas.

Conforme Gross (2024), Foucault propõe que o poder se manifesta via discursos que determinam normas e exclusões em diversas esferas sociais. Na perspectiva foucaultiana, a construção de verdades e a formação de subjetividades estão diretamente ligadas às práticas discursivas que reforçam hierarquias, neste caso do masculino frente ao feminino cis e principalmente em desfavor do trans, pela dominação dos bancos de dados desde a sua formação aliado com a não representação das novas identidades.

Pela não reformulação e pela não identificação da máquina em entender as subjetividades de cada sujeito e suas mudanças no transcorrer dos tempos e no espaço, sendo a máquina uma figura

Revista ARACÊ
ISSN: 2358-2472

estática onde, para alterar os dados precisa de auditorias e controles de qualidade humana, na mesma medida que mudam as normas sociais de inclusão de pessoas na atualidade. Todavia, não é isso que se percebe no cotidiano das máquinas, sendo uma estrutura de exclusão em que se perpetua banco de dados masculinos e heterossexuais, sem inserção e remodelação contínua de dados que traga ao mesmo direitos iguais entre o sujeito masculino e o feminino, com prevalência do corpo heterossexual frente as novas expressões de identidades de gênero, que pouco faz para mudar este sistema (Repo, 2015).

4 RESULTADOS

4.1 PROPOSTAS DE SOLUÇÕES PRÁTICAS E RECOMENDAÇÕES PARA INTOLERÂNCIA ALGORÍTMICA

Aplicando esse conceito aos algoritmos de IA, é possível perceber como a linguagem utilizada pode ser instrumentalizada para manter sistemas opressores que marginalizam indivíduos trans, ao impor rótulos e estigmatizações. Pierre Bourdieu, conforme De Toni Junior (2024), por sua vez, introduz o conceito de *habitus* para descrever as disposições e práticas socialmente adquiridas que influenciam comportamentos e percepções. O *habitus* das máquinas com banco de dados que privilegia o natural e o convencional funciona como um mecanismo de reprodução das estruturas sociais vigentes, onde a linguagem e os símbolos são utilizados para reforçar ou desafiar relações de poder.

Sua aplicação nas análises de sistemas de IA permite identificar como certas práticas linguísticas podem excluir ou deslegitimar a identidade trans, contribuindo para a reprodução de preconceitos. A construção de algoritmos de IA envolve a coleta e processamento de grandes volumes de dados, os quais refletem frequentemente as desigualdades e os estereótipos existentes na sociedade. Estudos indicam que modelos de linguagem podem associar determinadas profissões ou características a gêneros específicos, perpetuando estigmas (Araújo e Araújo, 2024). Essa replicação de preconceitos se torna ainda mais preocupante quando aplicada a contextos sensíveis, como a saúde, educação e emprego para a comunidade trans.

A resistência inclusiva na linguística aplicada às tecnologias de IA se manifesta, sobretudo, por meio do reconhecimento e valorização da diversidade do falar, ou seja, das múltiplas formas de expressão que caracterizam a esfera linguística dos indivíduos trans. Conforme apontado por estudos de Araújo e Araújo (2024), a implementação de linguagens que respeitem a identidade de gênero e as especificidades culturais são um passo determinante para a promoção de ambientes mais inclusivos.

A luta por reconhecimento e a resistência contra a exclusão se traduziam, na prática, em movimentos que desafiam as coisas como são e exigem a reformulação de práticas discriminatórias,

Revista ARACÊ
ISSN: 2358-2472

tanto no âmbito tecnológico quanto nas políticas públicas. A perspectiva em Foucault (1990) nos permite compreender como a construção do discurso, frequentemente pautada por estruturas de poder e exclusão, pode e deve ser transformada; ao mesmo tempo, a teoria do *habitus*[3] de Bourdieu (1993) evidencia que tais práticas são internalizadas e, por isso, é necessário um esforço contínuo para a desconstrução de preconceitos incorporados nos sistemas de IA.

Para compreender melhor, o conceito de *dispositivos de poder* refere-se a uma rede de práticas e instituições que influenciam o comportamento e as identidades das pessoas, enquanto biopoder diz respeito ao controle que sociedades exercem sobre a vida e o corpo dos indivíduos.

Assim, os dispositivos de poder, conforme Veyne (2024), moldam a subjetividade humana, operam tanto na história quanto na modernidade, criando um panorama onde certas identidades são exaltadas enquanto outras são marginalizadas. Ao explorar essa dinâmica por meio das relações do sujeito com o algoritmo, podemos perceber como a luta por reconhecimento e a contestação das normas sociais sempre existiram como um fio condutor na experiência humana.

A aplicação da inteligência artificial na saúde e no banco de dados de empregabilidadepossui um potencial grande tanto para promover a inclusão quanto para reforçar a marginalização, dependendo da qualidade dos dados e da sensibilidade das linguagens empregadas. Na área da saúde, a utilização de sistemas automatizados para diagnósticos ou atendimento pode resultar em práticas imprecisas, ou preconceituosas, caso os algoritmos não sejam treinados com dados representativos das especificidades da comunidade trans (Frazão, 2021).

Em contextos educacionais, a ausência de uma representação adequada das identidades e particularidades da comunidade trans pode levar a práticas pedagógicas excludentes, onde o viés linguístico influencia tanto a forma de ensino quanto a interação entre educadores e discentes (Hashiguti e Fagundes, 2022). Assim, a adoção de linguagens inclusivas e a revisão crítica dos dados utilizados são essenciais para construir práticas que respeitem e valorizem a diversidade.

A interseção de vieses linguísticos e algoritmos de IA, como discutido por Freitag (2024), revela que a falta de uma abordagem inclusiva pode resultar em erros sistêmicos, deteriorando a qualidade dos serviços prestados e potencializando o ostracismo e a discriminação contra indivíduos trans. Portanto, a integração dos estudos linguísticos na formação e validação dos algoritmos é fundamental para a promoção de um ambiente de atendimento que seja equitativo e sensível às necessidades de todos.

[3] *Habitus* em Bourdieu, significa as siposições internas que orientam a perepção do pensamento da pessoa, bem como a ação dos indivíduos. Pode ser considerada uma “segunda natureza” que se molda pelas experiências sociais, escolhas e comportamentos dos indivíduos.

Revista ARACÊ
ISSN: 2358-2472

Essa constatação sublinha a importância de combinar técnicas de *debiasing* com a supervisão humana, que permite uma interpretação crítica dos resultados e a aplicação de medidas corretivas adicionais. No setor de saúde, a implementação de métodos de *debiasing* enfrenta desafios similares. Dados clínicos, por vezes, refletem desigualdades históricas na forma de diagnósticos e tratamentos, o que pode resultar em modelos que não são suficientemente sensíveis às especificidades de gênero.

Para Matias e De Morais Junior (2024), tendo em vista os desafios expostos tanto nos sistemas de emprego quanto na área de saúde, é imperativo que políticas públicas e estratégias institucionais sejam repensadas para promover a igualdade de gênero em ambientes automatizados. Dentre as soluções potenciais, podemos destacar propostas em sistemas de auditorias regulares de gênero, onde a realização periódica nos algoritmos utilizados por instituições brasileiras é essencial para identificar possíveis desvios e corrigir vieses persistentes. Ferramentas como o *Aequitas*[4] podem ser empregadas para testar modelos em múltiplas métricas de equidade e viés (Saleiro *et al*., 2018). Gestores de IA devem ser instados a adotar essas práticas, assegurando a prestação de contas e a transparência dos processos.

A promoção de equipes diversificadas na composição dos times de desenvolvimento desempenha papel importante na construção de sistemas justos. A inclusão de mulheres cis e principalmente trans e profissionais de diferentes origens culturais e sociais é indispensável para evitar que algoritmos não reflitam somente uma perspectiva homogênea e, consequentemente, discriminatória.

A intervenção humana e a supervisão contínua continuam a ser um componente crítico na interpretação dos resultados obtidos por algoritmos. Especialistas em ciências humanas e ética devem fazer parte dos processos de revisão para garantir que os resultados automáticos sejam alinhados aos princípios de não discriminação.

A transparência de prestação de contas na abertura dos processos algorítmicos e o direito à explicação de decisões devem ser garantidos, para quequalquer possível exclusão possa ser imediatamente identificada e retificada. A LGPD oferece um arcabouço legal que, se bem aplicado, pode limitar e reduzir os riscos de discriminação conforme (Santos e Graminho, 2024). A atualização e qualidade dos dados é fundamental para que os bancos de dados utilizados para treinamento de algoritmos sejam constantemente atualizados e representem, equilibradamente, a diversidade dos públicos atendidos, seja no setor de emprego ou de saúde.

[4] Sistema que visa promover a igualdade de gênero nos processos informacionais e IA nas empresas. No segmento de emprego, visa modificar processos de contratação e remuneração, para serem o mais próximo possível da igualdade entre homens e mulheres.

Revista ARACÊ
ISSN: 2358-2472

A questão da transparência, para Zuiderveen Borgesius (2018), cumpre um papel fundamental na avaliação do impacto dos métodos de *debiasing*. Sistemas com decisões opacas dificultam a identificação e correção de vieses, e a transparência é apontada como uma solução que, aliada à responsabilização, pode promover a confiança dos usuários e a integridade dos processos. Essa necessidade de transparência nos algoritmos é reconhecida pela literatura e reforçada pelos princípios da LGPD, que preconiza a prestação de contas e o direito à explicação dos processos decisórios.

Essas propostas envolvem um compromisso coordenado entre desenvolvedores de sistemas de IA, formuladores de políticas e gestores institucionais. A integração dessas medidas não somente contribuirá para a redução dos vieses de gênero, mas também promoverá um ambiente digital mais ético e equitativo, alinhado aos princípios da justiça social.

Cabe destacar que a aplicação de técnicas de *debiasing*, embora valiosa, não pode ser vista como uma solução autônoma. A complexidade das relações sociais e a pluralidade das experiências de gênero demandam uma abordagem que considere tanto aspectos tecnológicos quanto as nuances culturais e históricas afetadas por esses sistemas automatizados.

Os resultados obtidos a partir da análise combinada do *corpus* linguístico e dos métodos de *debiasing* revelam padrões significativos de discriminação. Primeiramente, constatou-se que os algoritmos geradores de imagem reproduzem estereótipos visuais que reforçam normas binárias (divisões entre masculino e feminino). De acordo com Salmoria e Ramos (2024), esses sistemas tendem a limitar a representação de pessoas trans a papéis estereotipados, associando-as a imagens que perpetuam preconceitos.

Em paralelo, os algoritmos de processamento de linguagem natural demonstraram vieses sutis, mas sistemáticos. A análise dos *embeddings* de palavras mostrou que termos associados a identidades transgêneras e não binárias frequentemente aparecem em contextos negativos ou marginalizantes. Como constatado por Gonen e Goldberg (2019), os métodos de *debiasing* utilizados, embora capazes de reduzir parte do viés, não eliminam essas associações discriminatórias.

Outro resultado importante diz respeito à influência desses vieses em sistemas que operam em cenários de contratação e saúde. Conforme exposto por De Lima Viana (2023), os algoritmos utilizados em processos seletivos e na avaliação de necessidades médicas tendem a priorizar candidatos e pacientes que se enquadram em categorias de gênero tradicionalmente aceitas, evidenciando a dimensão prática e as consequências reais da discriminação algorítmica.

A análise dos dados revelou que as técnicas de *debiasing* aplicadas ao *corpus* analisado apresentam eficácia limitada. Embora haja uma redução nos vieses, o conjunto de dados ainda retém

Revista ARACÊ
ISSN: 2358-2472

traços de preconceitos históricos e culturais, corroborando a conclusão de que os métodos atuais são apenas paliativos e não oferecem uma solução definitiva (Gonen e Goldberg, 2019).

Os dados sugerem que a adoção de uma abordagem crítica que considere tanto aspectos linguísticos quanto contextuais pode contribuir para o desenvolvimento de tecnologias mais inclusivas. Os resultados indicam que a integração de metodologias da linguística crítica nos processos de treinamento de IA representa uma estratégia promissora para mitigar as disparidades identificadas. Um dos pontos centrais da discussão é a necessidade de reformular o treinamento e a construção dos algoritmos. A partir do debate promovido por Taso *et al*. (2023), fica claro que a inclusão de uma perspectiva crítica, que questione e desconstrua preconceitos históricos, é essencial para o desenvolvimento de sistemas de IA que respeitem a diversidade. Essa discussão também se estende à prática, onde desenvolvedores e projetistas de IA devem incorporar mecanismos que permitam a revisão contínua dos dados utilizados no treinamento dos modelos, revistando-os de forma rotineira.

Sobre às limitações dos métodos de *debiasing,* embora estudos como os de Gonen e Goldberg (2019) mostrem a eficácia parcial dessas abordagens, a persistência dos vieses com tendências que distorcem a objetividade indica que soluções tecnológicas isoladas não são suficientes para erradicar o preconceito. É necessário um esforço conjunto que envolva, além dos avanços técnicos, uma reavaliação dos pressupostos culturais e sociais que sustentam a construção dos dados.

A investigação dos impactos da discriminação algorítmica, preconceitos ou desigualdades resultantes do uso de algoritmos, conjuntos de regras para resolver problemas em áreas sensíveis, como saúde, educação e emprego, levanta questões éticas e políticas de grande relevância. Conforme destacado por De Lima Viana (2023), a exclusão de pessoas trans e não binárias por meio de sistemas de IA não se limita a um problema técnico; trata-se de uma questão que reflete exclusão social, exigindo intervenção governamental e regulatória.

A discussão se amplia para a necessidade de políticas públicas que façam as empresas desenvolvedoras de IA adotar práticas mais transparentes e inclusivas para poder haver a integração entre a linguística crítica e o desenvolvimento tecnológico, que se apresenta como uma estratégia inovadora para identificar e corrigir lacunas no tratamento das identidades diversas nos sistemas de IA.

Ao promover uma análise que combina métodos quantitativos e qualitativos, o estudo enfatiza a importância de um olhar atento às consequências sociais da automação e do processamento de linguagem. Essa integração pode, em última análise, contribuir para um ambiente digital mais justo e representativo.

Revista ARACÊ
ISSN: 2358-2472

4.2 SINERGIA ENTRE INTELIGÊNCIA ARTIFICIAL E SUJEITO HUMANO

Para Yu (2024), apesar do avanço tecnológico, a eficácia dos métodos de *debiasing* não pode ser alcançada exclusivamente por meio da automação. A integração do conhecimento humano é determinante para a interpretação correta dos dados e para a adaptação dos algoritmos a contextos específicos. O sujeito humano, tanto na condição de especialista quanto na de usuário final, possui a capacidade de identificar nuances contextuais e culturais que escapam muitas vezes à análise automatizada.

Essa sinergia propicia avanços na técnica da IA para o oferece o processamento em larga escala e a identificação rápida de padrões, enquanto o julgamento humano valida, interpreta e ajusta os modelos com base em conhecimentos teóricos e práticos. Por exemplo, na revisão de textos educacionais, a experiência pedagógica permite discernir quando uma expressão regional é apenas representativa de diversidade linguística ou quando ela indica uma obliquidade que precisa ser mitigada.

A colaboração entre IA e sujeito humano é fundamental para a ética na implementação de sistemas automatizados. Conforme enfatizado por Caliskan e Brison (2017), a remoção de vieses algorítmicos requer uma combinação de esforços que vai desde a criação de conjuntos de dados diversificados até a implementação de auditorias regulares e readaptação nos processos de regulamentação nos sistemas em funcionamento. Os especialistas em linguística computacional, psicologia social e áreas correlatas devem atuar lado a lado com desenvolvedores de IA para garantir que os resultados produzidos sejam refinados e inclusivos.

A participação humana não se limita somente à fase de treinamento dos algoritmos. Nos processos decisórios, o *feedback* contínuo e a revisão dos resultados pelos usuários e especialistas garantem a correção de possíveis desvios que possam surgir após a implementação dos sistemas. Assim, a união entre IA e sujeito humano se estabelece como a base para um modelo de *debiasing* sustentável e progressivo.

Em políticas públicas, essa integração pode ser incentivada entre parcerias através da globalização da informação no Brasil e no mundo onde as pessoas estão interligadas *on-line* que envolvem universidades, institutos de pesquisa e órgãos governamentais, incentivando projetos conjuntos que visem ao aprimoramento dos sistemas de tomada de decisão. A transparência nos processos de análise dos dados e a prestação de contas são pilares fundamentais para a construção de uma sociedade mais justa e equitativa.

Estudos na área de recrutamento evidenciam que algoritmos podem favorecer candidatos que se enquadram em padrões linguísticos previamente definidos como “ideais”, desconsiderando a

Revista ARACÊ
ISSN: 2358-2472

diversidade linguística e cultural. Técnicas de *debiasing* para sistemas de recrutamento incluem a revisão dos descritores textuais e a implementação de mecanismos de auditoria, onde especialistas em Recursos Humanos (RH) e cientistas de dados colaboram para ajustar os parâmetros dos modelos.

Importante é a conscientização sobre a necessidade de um monitoramento contínuo dos bancos de dados e dos algoritmos que os utilizam, uma vez que os vieses podem emergir ou se transformar com o tempo.

No contexto hospitalar, o *debiasing* implica na revisão dos relatórios clínicos e na implementação de sistemas de consulta que reconheçam variações linguísticas sem atribuir julgamentos subjetivos. O uso de técnicas de processamento de linguagem natural para analisar o contexto semântico pode ajudar a identificar expressões ambíguas e corrigi-las automaticamente. Ademais, a participação de profissionais de saúde e linguistas na validação dos dados garante que as intervenções sejam mais precisas e culturalmente sensíveis.

Em ambientes educacionais, os bancos de dados costumam englobar informações sobre desempenho acadêmico, participação estudantil e *feedback* de professores. Dados textuais provenientes de avaliações, fóruns de discussão e interações em plataformas *online* podem conter termos e expressões que refletem preconceitos implícitos que marginalizam certos grupos socioeconômicos e culturais. Por exemplo, a utilização de uma linguagem regional ou de gírias pode ser interpretada equivocadamente como indicativo de menor capacidade acadêmica, perpetuando estigmas e desigualdades similares à questão do gênero (Bolukbasi *et al.*, 2016).

A aplicação de técnicas de *debiasing* neste contexto requer uma análise detalhada dos padrões linguísticos e a identificação dos termos que possam ser associados a preconceitos. Medidas como a padronização dos textos, o uso de algoritmos capazes de sugerir correções linguísticas e a supervisão de especialistas em educação podem auxiliar na redução desses vieses. A colaboração entre áreas multidisciplinares, que integram conhecimentos da linguística computacional e da pedagogia, é fundamental para que o banco de dados reflita a diversidade sem comprometer a imparcialidade dos resultados.

5 CONCLUSÃO

A análise apresentada evidencia a importância dos estudos linguísticos na redução dos preconceitos presentes em algoritmos de IA que atuam em áreas cruciais, como a saúde, educação e emprego para a comunidade trans. Os conceitos teóricos de Foucault e Bourdieu oferecem uma base histórica e crítica para compreender como discursos e práticas sociais podem construir estereótipos e a marginalização de grupos historicamente oprimidos.

Revista ARACÊ
ISSN: 2358-2472

A implementação de medidas que garantam a revisão crítica da linguagem utilizada pelos algoritmos e a inclusão de dados representativos desponta como uma estratégia para mitigar os vieses algorítmicos. Ademais, a promoção de equipes diversificadas e a formação de parcerias interdisciplinares podem conduzir ao desenvolvimento de uma inteligência artificial que respeite e celebre a pluralidade das identidades.

O trabalho demonstrou que o preconceito de gênero na linguística crítica, quando transposto para o campo da IA, resulta em mecanismos de discriminação que impactam significativamente pessoas transgêneras e não binárias. Através da análise de conjuntos de textos específicos e da aplicação de métodos de *debiasing* foi possível identificar a persistência de vieses que reforçam normas tradicionais e excludentes.

Os resultados ressaltam a necessidade de repensar o processo de construção e treinamento dos algoritmos de IA, integrando práticas que considerem a diversidade e a pluralidade das identidades de gênero. A incorporação dos conhecimentos da linguística crítica proporciona uma perspectiva valiosa para a compreensão dos mecanismos de exclusão e, assim, se torna fundamental para que desenvolvedores, acadêmicos e ativistas possam colaborar na promoção de tecnologias verdadeiramente inclusivas destacam (Gontijo e Vogel, 2024).

Embora os métodos de *debiasing* mostrem progresso, eles ainda não conseguem eliminar os preconceitos historicamente enraizados nos dados. Portanto, a adoção de abordagens interdisciplinares, que combinem análises críticas com inovações tecnológicas, é indispensável para construir uma IA que respeite e celebre a diversidade.

Espera-se que as discussões e resultados apresentados neste artigo sirvam como base para futuras pesquisas e intervenções tecnológicas destinadas à neutralização dos preconceitos de gênero. Por meio do diálogo, na análise da linguagem e seus impactos sociais e no desenvolvimento tecnológico, pode-se imaginar um futuro nos quais a inteligência artificial atue como aliada na promoção da justiça social e na valorização da pluralidade humana.

Pesquisadores e formuladores de políticas públicas-institucionais devem se engajar ativamente na construção de um arcabouço regulatório que enfatize não somente a eficiência dos sistemas, mas também a justiça e a equidade nas suas aplicações.

Os estudos alertam para que as lacunas na representatividade dos dados podem levar a decisões que reforçam desigualdades históricas e perpetuam a exclusão dos grupos minoritários.

Para a redução do preconceito de gênero em algoritmos, envolve a integração de práticas tecnológicas, éticas e administrativas. Uma maior transparência, a diversidade nas equipes e a responsabilidade na gestão dos sistemas computacionais podem transformar os processos decisórios,

Revista ARACÊ
ISSN: 2358-2472

promovendo um ambiente mais inclusivo na área de emprego e saúde, principalmente. Tal postura deve ser considerada uma opção, mas uma necessidade diante dos desafios colocados pelos sistemas automatizados, além da sinergia entre o humano e a IA para reduzir os preconceitos.

Revista ARACÊ
ISSN: 2358-2472

REFERÊNCIAS

ARAÚJO, Júlio; DE ARAÚJO, Júlio Cézar Dantas. Racismo algorítmico e inteligência artificial: uma análise crítica multimodal. Revista Linguagem em Foco, v. 16, n. 2, p. 89–109. 2024.

BOURDIEU, Pierre et al. (Ed.). Bourdieu: critical perspectives. University of Chicago Press. 1993.

CALISKAN, Aylin; BRYSON, Joanna J.; NARAYANAN, Arvind. Semantics derived automatically from language corpora contain human-like biases. Science, v. 356, n. 6334, p. 183–186. 2017

DE TONI JUNIOR, Claudio Noel. Dispositivos de dominações: O capital biopolítica de dominações. Humanidades & Inovação, v. 11, n. 3, p. 227–240. 2024.

DEV, Sunipa et al. Harms of gender exclusivity and challenges in non-binary representation in language technologies. arXiv preprint arXiv:2108.12084. 2021.

GONEN, Hila; GOLDBERG, Yoav. Lipstick on a pig: Debiasing methods cover up systematic gender biases in word embeddings but do not remove them. arXiv preprint arXiv:1903.03862. 2019.

LIMA VIANA, Guilherme Manoel. Desigualdade na era digital: Como a discriminação algorítmica afeta os transexuais. Inova Jur., v. 2, n. 1, 2023.

FERNANDES, Erika Ribeiro; GRAGLIA, Marcelo Augusto Vieira. Inteligência Humana e Inteligência Artificial e os desafios dos vieses nos algoritmos de IA. RISUS–Journal on Innovation and Sustainability, São Paulo, v. 15, n. 1, p. 133–142, 2024.

FERNANDES, Dora Lorejan Avila et al. Promoção de Igualdade de Gênero: O uso de Inteligência Artificial em Processos de Recrutamento e Seleção em Empresas Brasileiras. FGV RIC Revista de Iniciação Científica, v. 3. 2022.

FOUCAULT, Michel. Qu'est-ce que la Critique?Bulletin de la Société Française de Philosophie, Paris, t. LXXXIV, année 84, n.2, p.35–63, avr./juin. 1990.

FREITAG, Raquel Meister Ko. Diversidade Linguística e inclusão ao digital: desafios para uma ia brasileira. arXiv preprint arXiv:2411,01259, 2024.

GONTIJO, Danielly Cristina Araújo; VOGEL, Stefani Juliana. Desafios de gênero na tecnologia e políticas para a inclusão. . In: IV Congresso internacional de direitos humanos de Coimbra: Uma visão transdiciplinar. p. 93. 2024.

GROS, Frédéric. A parrhesia em Foucault (1982–1984). GROS, Fréderic et al. 2004.

HASHIGUTI, Simone; FAGUNDES, Isabella Zaiden Zara. O algoritmo como materialidade discursiva em um contexto de educação Linguística. Letras & letras. 2022.

KOTEK, Hadas; DOCKUM, Rikker; SUN, David. Gender bias and stereotypes in large language models. In: Proceedings of the ACM collective intelligence conference, p. 12–24. 2023.

Revista ARACÊ
ISSN: 2358-2472

MATIAS, João Luís Nogueira; DE MORAIS JÚNIOR, Ricardo Antônio Maia. Discriminação algorítmica na relação de emprego: eficiência econômica, inteligência artificial e fragilidade do empregado. Revista do Tribunal Superior do Trabalho, v. 90, n. 2, p. 128–147. 2024.

O'NEIL, Cathy. Weapons of math destruction. How big data increases, inequality and threatens democracy. Crown, 2017.

RAFAELA SILVA, Mariah. Orbitando telas… Sur: International Journal on Human Rights/Revista Internacional de Direitos Humanos, v. 18, n. 31, 2021.

REPO, Jemima. The biopolitics of gender. Oxford University Press. 2015.

SALEIRO, Pedro et al. Aequitas: A bias and fairness audit toolkit. arXiv preprint arXiv: 1811.05577 2018.

FRAZÃO, Ana. Transparência de algoritmos x segredo de empresa. 2021.

SALMORIA, Camila Henning; RAMOS, Wanessa Assunção. Discriminação algorítimica: desvendando a representação de pessoas trans e queer por modelos de IA geradores de imagem. Revista CNJ, Brasília, v. 8, n. 2, p. 211–226. 2024. DOI: 10.54829/revistacnj.v8i2.618. Disponível em: https://www.cnj.jus.br/ojs/revista-cnj/article/view/618. Acesso em: 9 jul. 2025.

SANTOS, Rodrigo Coimbra; GRAMINHO, Vivian Maria Caxambu. Discriminação algorítmica nas relações de trabalho e princípios da Lei Geral De Proteção De Dados. Sequência (Florianópolis), v. 45, n. 96, p. e96294. 2024.

TASO, Fernanda Tiemi de S.; REIS, Valéria Q.; MARTINEZ, Fábio HV. Discriminação algorítmica de gênero: Estudo de caso e análise no contexto brasileiro. In: Workshop sobre as Implicações da Computação na Sociedade (WICS). SBC, 2023. p. 13–25. 2023.

VEYNE, Paul. Foucault: seu pensamento, sua pessoa. Civilização Brasileira. 2024.
YU, Chen. Gender Inequality in the Age of AI: Predictions, Perspectives, and Policy Recommendations. Center for Open Science. 2024.

ZUIDERVEEN BORGESIUS, Frederik et al. Discrimination, artificial intelligence, and algorithmic decision-making. Council of Europe, Directorate General of Democracy, v. 42. 2018.