1. Introdução

Nas últimas décadas, a Inteligência Artificial (IA) tem se consolidado como uma das tecnologias mais influentes no cotidiano contemporâneo, com aplicações que vão desde a saúde e segurança pública até sistemas de recomendação em redes sociais e produção de conteúdo textual. Apesar de seus benefícios, estudos recentes têm evidenciado um problema grave: a reprodução de viéses sociais e raciais, sobretudo contra pessoas negras, nos resultados gerados por sistemas de IA. Modelos de linguagem treinados a partir de grandes volumes de dados textuais, retirados da internet e repletos de preconceitos estruturais, acabam por reforçar estereótipos já presentes na sociedade como a associação da negritude a traços negativos ou subalternizados.

2. Problema

Esse problema se agrava com o uso de técnicas de aprendizado profundo (deep learning), que tornam o processo de tomada de decisão da IA mais opaco e difícil de interpretar. Conforme apontado no artigo *Da "Caixa-Preta" à "Caixa de Vidro"*, a ausência de explicabilidade (transparência) nos modelos algorítmicos limita a identificação e correção desses vieses, tornando urgente a discussão sobre o uso responsável e ético da tecnologia. Assim, justifica-se a necessidade de desenvolver pesquisas que compreendam e enfrentem o racismo algorítmico, ampliando o debate sobre justiça social no campo tecnológico.

3. Objetivo Geral

Este trabalho tem como objetivo geral analisar como os sistemas de IA baseados em dados textuais reproduzem preconceitos raciais contra pessoas negras.

4. Proposta de Solução

Como proposta de solução, propõe-se refletir sobre mecanismos técnicos e éticos que tornem os sistemas mais transparentes, com foco especial na análise da Inteligência Artificial Explicável (XAI) como estratégia para mitigar esses efeitos.

5. Metodologia

A metodologia do trabalho é de natureza qualitativa e bibliográfica, com base na análise de artigos científicos, estudos de caso e relatos de pesquisadores que abordam o viés racial em modelos generativos de texto. Além disso, o estudo dialoga com contribuições das áreas da linguística, semiótica, estudos culturais e ética em tecnologia. Trechos produzidos por modelos de IA também serão analisados criticamente, buscando-se evidenciar padrões discriminatórios e associá-los aos dados de treinamento e decisões algorítmicas subjacentes.

6. Resultados Esperados

Como resultado, espera-se identificar como os preconceitos são codificados e reproduzidos pelas IAs textuais, apontando as falhas estruturais nos dados e modelos. O trabalho também buscará mapear os limites e as potencialidades da XAI na correção desses padrões discriminatórios.

7. Considerações Finais

Conclui-se que, para garantir um desenvolvimento tecnológico mais justo, é essencial considerar as implicações sociais e históricas que atravessam os dados e os algoritmos. A IA, enquanto ferramenta, deve ser acompanhada de mecanismos que promovam responsabilidade, diversidade e ética, especialmente quando seus impactos recaem sobre grupos historicamente marginalizados, como a população negra.

8. Referência

Silva, Tarcízio & Ribeiro, Rafael Evangelista. *Da “Caixa-Preta” à “Caixa de Vidro”: inteligência artificial, explicabilidade e justiça algorítmica*. InternetLab, 2021. Disponível em: https://www.internetlab.org.br (acesso em ago. 2025).