

RACISMO ALGORÍTMICO: UMA DISCUSSÃO FUNDAMENTADA

Pedro Martins de Oliveira Menezes¹; Lucas Amparo Barbosa²

¹ Graduando em Engenharia da Computação; pedro.menezes@aln.senaicimatec.edu.br

² Centro Universitário SENAI CIMATEC; Salvador - BA; lucas.barbosa@fieb.org.br

RESUMO

Modelos de Aprendizado de Máquina, subárea da Inteligência Artificial, vem se tornando cada vez mais comuns no cotidiano. Contudo, há a necessidade de analisar o impacto que esses modelos podem ter em minorias socialmente vulneráveis e os vieses que podem ser apresentados, independentemente de sua causa e mitigando sua consequência. Para isso, é importante entender incidentes que já ocorreram no passado, observando suas potenciais falhas técnicas e consequências jurídicas relacionadas. A partir dessa análise, o trabalho identifica e relaciona soluções apontadas nos trabalhos estudados que podem minimizar o impacto das falhas, principalmente quando relacionadas a minorias diversas.

PALAVRAS-CHAVE: Racismo Algorítmico; Aprendizado de Máquina; Visão Computacional; Inteligência Artificial.

1. INTRODUÇÃO

A Inteligência Artificial (IA) vem se tornando uma ferramenta relevante à vida em sociedade nos últimos anos. Grande parte dos aplicativos, sites e serviços que são utilizados atualmente contam com algum tipo de suporte inteligente, podendo qualquer desenvolvedor contratar esses tipos de recursos.¹ Na seara dessa evolução, há o destaque para os algoritmos que usam o aprendizado de máquina, do inglês *machine learning*, que trata do reconhecimento de padrões através de uma base de dados e posterior aplicação do aprendizado no reconhecimento das variáveis em outras unidades ou conjuntos de dados.² Em primeira instância, isso não parece ter nada de problemático ou perigoso, porém, ao perceber que o resultado de um processo de predição é função dos dados que lhe são fornecidos, surgem questionamentos: o que acontece se esses dados forem, propositalmente ou não, distorcidos? Esses modelos são passíveis de manipulação?

O presente artigo tem como propósito analisar a literatura acadêmica sobre racismo e discriminação algorítmica, a fim de trazer à tona as discussões levantadas, relacioná-las e discutir os problemas gerados por algoritmos que acabam vitimizamando minorias raciais e étnicas, e analisar *datasets* atuais em uso para, também, relacioná-los às discussões da literatura e validá-las, culminando em elencar possíveis soluções e avaliar sua relevância e tangibilidade.

2. METODOLOGIA

Os artigos para revisão bibliográfica foram buscados nas plataformas Google Acadêmico e Scielo, utilizando combinações de palavras-chave como “Racismo Algorítmico”, “Discriminação Algorítmica”, “Viés Algorítmico”, “Inteligência Artificial”, “Racismo”, e afins, assim como seus equivalentes em língua inglesa. Também foi usado um E-Book compartilhado pelo orientador do projeto como fonte. Após a pré-seleção e leitura dos artigos, seus temas levantados foram organizados numa tabela de metadados para a comparação com o resto da amostra e assim aferir sua relevância e determinar as relações existentes, assim selecionando os que seriam efetivamente usados como referencial teórico. Foram utilizados 9 artigos base, divididos entre aqueles que faziam análises tecnológicas e os que faziam análises jurídicas, a fim de organizar pensamentos e problemáticas consoantes.

Os *datasets* analisados foram selecionados na plataforma Kaggle, utilizando palavras-chave como “*Ethnicity*”, “*Police*”, “*Race*” e “*Face*”. Também foram usados *datasets* citados nos artigos estudados. Após essa pré-seleção, os atributos desses conjuntos de dados foram verificados, e sua relevância para as análises trazidas no artigo foi filtrada com base nesses, chegando ao conjunto final de dados utilizados para a validação de hipóteses e afirmações trazidas na literatura. Foram utilizados 6 conjuntos de dados, sendo 5 deles dados de fontes policiais e 1 de dados de rostos de pessoas.

3. FUNDAMENTAÇÃO TEÓRICA

Silva usa o conceito de branquitude para discutir a falsa objetividade de sistemas automatizados, a predominância de pessoas brancas como desenvolvedoras de tecnologias e o conceito de “caixa preta”:

dispositivos que funcionam em termos de entrada de informações e decisões e saída de resultados e operações que ocultam os modos pelos quais ciência e tecnologia são construídas. O autor também levanta os vieses de dados de treinamento como agravante da falsa objetividade.¹ Silva e Araújo fazem uma análise exploratória de casos de racismo algorítmico para argumentar que a presença de algoritmos na vida cotidiana se tornou uma ferramenta no chamado racismo estrutural-algorítmico.³

Amaral, Martins e Elesbão fazem uma pesquisa investigativa sobre a neutralidade de algoritmos de bancos de imagem muito utilizados, comprovando a hipótese de que o “neutro” seria interpretado como “branco” pelos sistemas.⁴ Carrera questiona a falta de interesse dos donos de grandes sistemas de mudá-los fundamentalmente uma vez apontados como problemáticos, também questionando se é sempre necessário existir um estopim para fazer mudanças significativas em seus produtos em vez de fazê-las voluntariamente. A autora também explica o funcionamento de bancos de imagens digitais para exemplificar como a ideia de objetividade é equivocada.⁵

Lucena disserta sobre o Policiamento Preditivo, uso de algoritmos para auxiliar a polícia a prevenir crimes, explicando seu funcionamento, apontando suas inconsistências e mostrando como se trata de mais um mecanismo de perpetuação de vieses discriminatórios usando a bandeira da neutralidade.⁶

Possa argumenta como algoritmos de reconhecimento facial servem para agravar a segregação racial do sistema carcerário brasileiro ao afetar pessoas negras desproporcionalmente, e como seu uso é “consentido” pela sociedade sob a justificativa de auxiliar na segurança pública.⁷ Pereira, também sobre o reconhecimento facial, traz o conceito de inimigo no direito penal para expandir como essa tecnologia e, ademais, as utilizadas em automatização de decisões judiciais afetam mais as pessoas negras.⁸ Costa e Kremer dizem que essa tecnologia, também, é uma ferramenta de controle estatal e que, em seu estado atual de desenvolvimento e treinamento, ela serve para continuar atingindo aqueles que já eram considerados inimigos antes, os que devem ser vigiados e perseguidos pelo Estado e sociedade.⁹

Fidalgo volta seu olhar para o uso de algoritmos nas relações trabalhistas, mostrando que além de preconceito racial, eles também são capazes de agir de forma machista, discriminando até mesmo mulheres grávidas. A autora também retoma as questões de diversidade dos meios de produção e treinamento das tecnologias.¹⁰

Os *datasets* policiais analisados são todos desbalanceados quanto a gênero, tendo desproporcionalmente mais homens cadastrados do que mulheres. Quanto à raça, os conjuntos referentes a prisões são desbalanceados com mais cadastros de “negros”,^{11; 12; 13} com um deles tendo números mais elevados de “hispanicos/latinos” em comparação aos outros conjuntos.¹² Os referentes a fatalidades cometidas por policiais tem mais cadastros de “brancos”, porém com números elevados de “negros” e “hispanicos/latinos” também atingidos.^{14; 15} O conjunto de dados faciais, diferente dos conjuntos policiais, é balanceado quanto a gênero, porém desbalanceado quanto a raça. Assim como nos conjuntos de fatalidades, “brancos” são a raça individual mais representada, porém, diferente de todos os outros *datasets*, “negros” compõem menos de 5% do conjunto.¹⁶ Com a falta de equilíbrio apontada para as raças, o problema levantado por Silva de vieses de dados de treinamento e, considerando a origem policial da maioria dos dados, os levantados por Possa e Pereira sobre a desproporcionalidade de alvos negros é ratificada.^{1; 7; 8} Ademais, a discriminação de gênero levantada por Fidalgo também é reafirmada diante do desbalanceamento de dados de gênero.¹⁰ Com base no estudado, é possível afirmar que as disparidades encontradas na análise dos *datasets* são representativas de um universo de dados que são constantemente utilizados em aplicações e sistemas diversos, apesar do trabalho ter sido executado em uma amostra pequena, composta por 6 (seis) bases de dados.

4. CONSIDERAÇÕES FINAIS

O trabalho apresentado busca trazer luz ao tema dos vieses que podem ocorrer no processo de construção, treinamento e utilização de modelos inteligentes. De forma muito homogênea, a base de trabalhos anteriores que discutiram o tema sempre converge no ponto da ideia vendida de que a tecnologia em si é neutra, incapaz de carregar nenhum tipo de preconceito. Conforme discutido, isso não é verdade por diversos fatores, como seu desenvolvimento e dados fornecidos.

A mais promissora solução apontada pelos autores está relacionada diretamente à diversidade: Diversidade nos dados utilizados para construir e treinar os modelos; Diversidade nas equipes de programadores; Diversidade na difusão desses modelos e nos locais de utilização dos mesmos. Quanto mais diversos forem os ambientes, menor é a possibilidade de um viés problemático passar sem ser percebido.

Como trabalhos futuros relacionados a essa pesquisa, um processo de validação de potenciais soluções e estratégias de mitigação levantadas na literatura é extremamente pertinente. É possível criar um ambiente controlado e forçar os erros que já ocorreram antes e, assim, validar se a solução apontada de fato minimiza o impacto do viés naquele contexto. É coerente, também, a criação de diretrizes de desenvolvimento

que visem diminuir a criação de vieses nos modelos matemáticos, podendo fazer parte das respectivas legislações.

5. REFERÊNCIAS

- ¹ SILVA, Tarcísio da. **Visão computacional e racismo algorítmico**: Branquitude e opacidade no aprendizado de máquina. Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN), v. 12, n. 31, fev. 2020. Disponível em: <<https://abpnrevista.org.br/site/article/view/744>>.
- ² OLIVEIRA, Carla. **Aprendizado de máquina e modulação do comportamento humano**. In: SOUZA, J.; ALVELINO, R.; SILVEIRA, S. A. A Sociedade de Controle: manipulação e modulação nas redes digitais. São Paulo: Hedra, 2018.
- ³ SILVA, Mozart Linhares da; ARAÚJO, Willian Fernandes. **Biopolítica, racismo estrutural-algorítmico e subjetividade**. Revista Educação Unisinos, v. 24, out. 2020. Disponível em: <<https://revistas.unisinos.br/index.php/educacao/article/view/edu.2020.241.40>>.
- ⁴ AMARAL, Augusto Jobim do; MARTINS, Fernanda; ELESBÃO, Ana Clara. **Racismo algorítmico**: uma análise da branquitude nos bancos de imagens digitais. Pensar : Revista de Ciências Jurídicas, v. 26, n. 4, nov. 2021. Disponível em: <<https://ojs.unifor.br/rpen/article/view/11806>>.
- ⁵ CARRERA, Fernanda. **Racismo e sexismo em bancos de imagens digitais**: análise de resultados de busca e atribuição de relevância na dimensão financeira/profissional. In: TARCÍZIO SILVA. Comunidades, algoritmos e ativismos digitais: olhares afrodiaspóricos. São Paulo: Literarua, 2020. p. 149–165.
- ⁶ LUCENA, Pedro Arthur Capelari de. **Policamento preditivo, discriminação algorítmica e racismo**: potencialidades e reflexos no brasil. In: VI SIMPÓSIO INTERNACIONAL LAVITS: “ASSIMETRIAS E (IN)VISIBILIDADES: VIGILÂNCIA, GÊNERO E RAÇA”, VI., 2019, Salvador. Anais do VI Simpósio Internacional LAVITS: “Assimetrias e (In)Visibilidades: Vigilância, Gênero e Raça”. Salvador: Lavits, 2019. Disponível em: <<https://lavits.org/anais-do-vi-simposio-internacional-lavits-assimetrias-e-invisibilidades-vigilancia-genero-e-raca/>>.
- ⁷ POSSA, Alisson. **O reconhecimento facial como instrumento de reforço do estado de coisas inconstitucionais no Brasil**. IDP Law Review, v. 1, n.2, mar. 2022. Disponível em: <<https://www.portaldeperiodicos.idp.edu.br/lawreview/article/view/5943>>.
- ⁸ PEREIRA, Débora Freitas Mendes. **O uso de câmeras de reconhecimento facial em contexto de pós democracia – uma ferramenta contra o inimigo no direito penal?** ADPEB, 2020. Disponível em: <<http://www.adpeb.com.br/v18/wp/wp-content/uploads/2021/02/O-USO-DE-CÂMERAS-DE-RECONHECIMENTO-FACIAL-EM-CONTEXTO-DE-PÓS-DEMOCRACIA---UMA-FERRAMENTA-CONTRA-O-INIMIGO-NO-DIREITO-PENAL.pdf>>.
- ⁹ COSTA, Ramon Silva; KREMER, Bianca. **Inteligência artificial e discriminação**: desafios e perspectivas para a proteção de grupos vulneráveis frente às tecnologias de reconhecimento facial. Revista Brasileira de Direitos Fundamentais & Justiça, v. 16, n. 1, out. 2022. Disponível em: <<https://dfj.emnuvens.com.br/dfj/article/view/1316>>.
- ¹⁰ FIDALGO, Luiza Barreto Braga. **Discriminações algorítmicas**: racismo e sexismo nas relações laborais. Brazilian Journal of Development, v. 8, n. 10, p. 67341–67354, Oct. 2022. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BRJD/article/view/53113>.
- ¹¹ ARRESTS in the City of Chicago (2014 - 2023): banco de dados preparado por Matt OP. Disponível em: <<https://www.kaggle.com/datasets/mattop/arrests-in-the-city-of-chicago-2014-2023>>. Acesso em 22 out. 2023.
- ¹² DALLAS Police Arrests: banco de dados preparado por JohnM. In: **Police Violence & Racial Equity - Part 2 of 2**. Disponível em: <<https://www.kaggle.com/datasets/jpmiller/police-violence-racial-equity>>. Acesso em 29 out. 2023.
- ¹³ NYPD_Arrests_Data__Historic_: banco de dados preparado por JohnM. In: **Police Violence & Racial Equity - Part 2 of 2**. Disponível em: <<https://www.kaggle.com/datasets/jpmiller/police-violence-racial-equity>>. Acesso em 27 out. 2023.
- ¹⁴ DATA Police shootings: banco de dados preparado por Andrii Samoshyn. Disponível em: <<https://www.kaggle.com/datasets/mrmorj/data-police-shootings>>. Acesso em 25 out. 2023.
- ¹⁵ POLICE Violence & Racial Equity - Part 1 of 2: banco de dados preparado por JohnM. Disponível em: <<https://www.kaggle.com/datasets/jpmiller/police-violence-in-the-us>>. Acesso em 29 out. 2023.
- ¹⁶ AGE, GENDER AND ETHNICITY (FACE DATA) CSV: banco de dados preparado por Nipun Arora. Disponível em: <<https://www.kaggle.com/datasets/nipunarora8/age-gender-and-ethnicity-face-data-csv>>. Acesso em 28 out. 2023.