LUMINA
Revista do Programa de Pós-graduação em Comunicação
Universidade Federal de Juiz de Fora / UFJF
e-ISSN 1981-4070

Algoritmos da branquitude:

vieses e representações racistas em sistemas de inteligência artificial

Thaise Marques de Lima[1], Vanessa Maria Gomes da Silva[2] e Fellipe Sá Brasileiro[3]

Resumo

Este artigo objetiva refletir sobre o regime da branquitude em sistemas de inteligência artificial a partir de casos de viéses algorítmicos baseados em representações racistas. Parte do método do estudo de caso aplicado a dois escândalos de viés algorítmico envolvendo o *Copilot* e o *Gemini*, *chatbots* com sistemas de inteligência artificial produzidos pelas empresas de tecnologia Microsoft e Google. Destaca que as práticas de *machine learning* dos sistemas de inteligência artificial apontados são operadas com base nas estruturas de representação da branquitude,na medida em que são capazes de gerar informações imagéticas que associam pessoas de pele negra a um imaginário social formado por estereótipos, uma suposta representação há muito perpetuada por um sistema colonial racista. Observa que as soluções encaminhadas pelas empresas de tecnologia diante dos problemas relatados se restringem aos ajustes insuficientes nos produtos finais em detrimento de intervenções estruturais eficazes. Conclui-se que as empresas de tecnologia responsáveis por inteligência artificial poderiam apresentar respostas resilientes diante do problema do racismo algorítmico ao elaborarem estratégias eficazes para o aprendizado de máquina fundamentadas em bases de dados decoloniais e diversas, assim também como em treinamentos com *feedbacks* humanos.

Palavras-chave

Inteligência artificial; Racismo algorítmico; Representação; Branquitude; Vieses.

[1]Mestranda no Programa de Pós-graduação em Comunicação e Culturas Midiáticas. E-mail: thaisem.lima98@gmail.com.
[2]Mestranda no Programa de Pós-graduação em Comunicação e Culturas Midiáticas. E-mail: vanessasilva@tutamail.com.
[3] Doutor em Ciência da Informação. Docente do Programa de Pós-Graduação em Comunicação e Culturas Midiáticas da UFPB. E-mail: fellipe.brasileiro@academico.ufpb.br.

LUMINA
Revista do Programa de Pós-graduação em Comunicação
Universidade Federal de Juiz de Fora / UFJF
e-ISSN 1981-4070

Whiteness algorithms:

biases and racist representations in artificial intelligence systems

Thaise Marques de Lima[1], Vanessa Maria Gomes da Silva[2] and Fellipe Sá Brasileiro[3]

Abstract

This article aims to reflect on the regime of whiteness in artificial intelligence systems based on cases of algorithmic bias based on racist representations. It uses the case study method applied to two algorithmic bias scandals involving *Copilot* and *Gemini*, chatbots with artificial intelligence systems produced by the technology companies Microsoft and Google. It points out that the machine learning practices of the artificial intelligence systems mentioned are operated based on the structures of representation of whiteness, insofar as they are capable of generating imaginary information that associates people with black skin with a social imaginary formed by stereotypes and a supposed representativeness that does not correspond to historical facts. It notes that the solutions put forward by technology companies in the face of the problems reported are limited to insufficient adjustments to the final products, to the detriment of effective structural interventions. It concludes that technology companies responsible for artificial intelligences can present resilient responses to the problem of algorithmic racism by devising effective machine learning strategies based on decolonial and diverse databases, as well as training with human feedback.

Keywords

Artificial intelligence; Algorithmic racism; Representation; Whiteness; Biases.

[1]Mestranda no Programa de Pós-graduação em Comunicação e Culturas Midiáticas. E-mail: thaisem.lima98@gmail.com.
[2]Mestranda no Programa de Pós-graduação em Comunicação e Culturas Midiáticas. E-mail: vanessasilva@tutamail.com.
[3] Doutor em Ciência da Informação. Docente do Programa de Pós-Graduação em Comunicação e Culturas Midiáticas da UFPB. E-mail: fellipe.brasileiro@academico.ufpb.br.

Introdução

O imaginário social a partir de Teves (1992) pode ser compreendido como uma estrutura complexa; uma rede de sentidos e significados que envolvem uma determinada sociedade. Teves aponta que o imaginário social é um conjunto de relações construídas socialmente, dando sentido à nossa realidade por meio de artifícios simbólicos, um sistema que envolve "formas discursivas, falas múltiplas: escrita, gestual, imagética, enfim, modos simbólicos de dizer o mundo" (Teves, 1992, p. 14). A autora defende que é "mediante esse imaginário que o grupo se identifica, estabelece suas trocas, distribui seus papéis sociais" (Teves, 1992, p. 25). Desse modo, marcos históricos e culturais influenciam essa estrutura de sentidos em que os indivíduos e grupos estão envolvidos, estabelecendo formas de percepção e compreensão legítimas.

Grupos com maior domínio das "múltiplas falas", em meio a disputas e convergências, estabelecem a legitimação de diversos aspectos sociais. Esse exercício de poder envolve as estratégias simbólicas, representacionais de um grupo sobre outro, estabelecidas para demonstrar superioridade/inferioridade, como no período em vigência do tráfico de pessoas negras. A representação desses corpos era marcada por um juízo de desumanização que justificava tal barbárie, uma ideologia de naturalização da escravização durante séculos. Deve-se observar que essa não foi a única estratégia de dominação colonial, mas historicamente, através de representações mediante estereótipos, perpetuou-se um imaginário social que reforça e sustenta as estruturas de poder e dominação baseadas no racismo.

Nesse campo, o imaginário estabelece uma construção de sentido capaz de firmar um discurso ideológico no qual a inferiorização do corpo negro se objetiva. Os estereótipos criados reforçam discursos hegemônicos nos quais frequentemente corpos negros são associados à degeneração, violência e dependência. Com efeito, a imprensa tradicional foi historicamente central no processo de reprodução desse imaginário racista enquanto "campo fértil de produção simbólica e também lugar de referência para a observação de imagens sociais" (Lopes; Lins, 2019, p. 92). Na mesma direção, a partir de lógicas diferentes, as tecnologias digitais de informação e comunicação que atualizaram os meios de comunicação tradicionais continuam a reproduzir uma ordem que reforça esse sistema de dominação.

A mediação dos algoritmos, em especial, se configura como a lógica complexa que reproduz o racismo na medida em que ocorre entrelaçada com as circunstâncias sociais, econômicas, culturais e políticas baseadas no racismo. Tal entendimento, presente nas reflexões de trabalhos recentes no campo da comunicação, é fundamental para a refutação de qualquer argumento que se formule a partir da ideia de neutralidade das tecnologias digitais.

Com base em Noble (2018), argumentamos que um dos maiores recursos para

compreender as estratégias da opressão algorítmica é a consciência de que essas formulações matemáticas, que guiam as decisões automatizadas, são feitas, em sua maioria, por seres humanos. Assim, no *back-end* de qualquer sistema computacional há interesses, verdades e princípios próprios. A utopia de uma sociedade digital igualitária – presente na declaração de independência do ciberespaço – tem se desmistificado cada vez mais diante de estudos e pesquisas aprofundadas que, por meio de dados reais, comprovam a presença de problemáticas raciais, de gênero e de classe (Noble, 2018; Benjamin, 2019; O'Neil, 2021).

É possível dizer, portanto, que os algoritmos e os sistemas de inteligência artificial que fazem a mediação de informações apresentam consideráveis contradições e controvérsias. De acordo com Noble (2018), ao compreender a "racialização tecnológica" como uma modalidade de opressão algorítmica, os pesquisadores têm acesso a um enquadramento conceitual para analisar de forma crítica os discursos que enaltecem a Internet como espaço democrático.

Nessa perspectiva, diferentes pesquisadores (Silva, 2020; 2022; Rodrigues *et al.*, 2023; Karam Tietboehl *et al.*, 2024) apontaram conexões entre os estudos sobre branquitude e as opressões algorítmicas, como o racismo algorítmico. O ponto em comum que liga tais esforços é o desenvolvimento de abordagens capazes de revelar como as tecnologias automatizadas continuam a mostrar vieses racistas, mesmo quando provenientes de grandes corporações globais com amplo suporte financeiro e tecnológico disponível (Silva, 2020).

Partindo de tais premissas, com o objetivo de compreender o regime da branquitude em sistemas de inteligência artificial baseados em representações racistas, realizamos um estudo de caso envolvendo dois escândalos de viés algorítmico. Concomitante a esse objetivo, buscamos refletir na direção da não neutralidade das tecnologias digitais, considerando as especificidades dos casos. Os dois casos analisados são de *chatbots* de inteligência artificial (IA) produzidos pelas empresas de tecnologia Google e Microsoft.

A estrutura do artigo reflete três objetivos específicos: (a) contextualizar casos de vieses algorítmicos em sistemas de inteligência artificial; (b) situar os mecanismos que conduzem a inteligência artificial a entregarem resultados enviesados e (c) analisar os escândalos sob a perspectiva de um imaginário social racista e os conceitos de branquitude.

A representação e suas implicações sociais

O modelo econômico escravista que perdurou ao longo de séculos é fundamentado em uma ideologia que justificou e legitimou o racismo. Essa ideologia foi massivamente difundida fazendo parte do imaginário social, um conjunto complexo de relações socialmente construídas para dar sentido à realidade, baseado

em elementos simbólicos como a "escrita, gestual, imagética" (Teves, 1992, p. 14). Essas estruturas simbólicas são importantes para compreensão do porquê o racismo existe de formas tão sutis. Em conformidade com Moura (2019), o racismo não é somente um resquício do sistema econômico mencionado, dado que as forças produtivas do capitalismo não o superam; trata-se de uma ideologia de dominação de um grupo sobre outro, em meio às evidentes estratégias de manter esse modelo de hierarquia estão as múltiplas formas de barragem da mobilidade social da população negra, após a abolição da escravatura.

Na formulação de Araújo (2019), o imaginário é um conjunto de relações que constituem um repertório, o "capital pensado" humano, por meio do qual se estabelece a construção de sentido individual e coletivo, sendo o imaginário social atravessado por complexidades culturais. Os grupos dominantes, em meio a disputas e convergências, com maior controle dos aparatos simbólicos, estabelecem sentido a diversos aspectos sociais, legitimando suas perspectivas de mundo. Esse exercício de poder envolve também as estratégias simbólicas, representacionais de um grupo sobre outro. Bento (2022) evidencia um importante aspecto desse poder, fruto da herança escravista acumulada de forma silenciosa e perpetuada pelo tempo:

> Descendentes de escravocratas e descendentes de escravizados lidam com heranças acumuladas em histórias de muita dor e violência, que se refletem na vida concreta e simbólica das gerações contemporâneas. Fala-se muito na herança da escravidão e nos seus impactos negativos para as populações negras, mas quase nunca se fala na herança escravocrata e nos seus impactos positivos para as pessoas brancas (Bento, 2022, p. 24).

O que direciona a maneira através da qual esses aparatos simbólicos, assim como econômicos, estão distribuídos de forma desigual entre os grupos sociais. As formas de representação que compõem um imaginário social referente ao lugar do negro são uma das estratégias de manutenção desses privilégios e revelam as circunstâncias nas quais a identidade negra está historicamente imergida. Desse modo, representações são indispensáveis na concepção de imagens sociais e suas significações, produzindo sentido, para construir realidades (Gomes, 2020, p. 78). Alinhada com Hall, Moraes (2019, p. 170) explica que "as representações têm sérias implicações sobre as identidades", que, por sua vez, são, de acordo com Gomes (2017, p. 41), "um modo de ser no mundo e com os outros".

Identidade seria um conjunto de características individuais que fazem sentido coletivamente – o sentimento de pertencimento que tem como recorte a diferença. De acordo com Moraes (2019), as identidades sociais são "pensadas e construídas no interior da representação, através da cultura, sendo resultantes de um processo de identificação" (Moraes, 2019, p. 170). Nesse sentido, é importante constatar que nas primeiras representações da população negra, ao que se refere às circulações massivas principalmente, marcada por uma lógica colonial que se impulsionou por

meio da imprensa, o negro "figurava como mercadoria", pois era onde se estampavam os "anúncios de compra, venda, aluguel e fuga de escravizados" (Gomes, 2020, p. 79). Nesse processo, o colonizador, ao representar "o outro", o faz o mais distinto possível de suas características, e "exagera as diferenças entre ele e o colonizado" (Sovik, 2020, p. 11) de modo a demarcar aspectos de superioridade e inferioridade, o preceito para a racionalização de um processo de exploração.

Nessa concepção, "o reconhecimento da diferença legitima a dominação" (Novaes, 1992, p. 125). Diante disso, é possível compreender como o racismo é algo além da "memória atávica da escravidão" (Sovik, 2020, p. 24); é marcado por um juízo de dominação, fundamentado na inferiorização da diferença. Assim, a pessoa negra tem sua identidade e história criada por outros (Lorena; Pio, 2022, p. 158), limitada, cercada por estereótipos nos quais "as relações de poder se expressam e se reafirmam" (Sovik, 2020, p. 6). A teórica afro-americana Patricia Hill Collins defende o conceito de "imagens de controle", ao categorizar estereótipos vinculados à mulher negra norte-americana, como uma forma de continuidade da dominação (Lorena; Pio, 2022).

A respeito do estereótipo, por definição, entende-se a partir do grego como "stereos, ou sólido, remonta à indústria gráfica e se refere às placas de metal fundido em moldes de gesso ou *papier-mâché*, que permitiam gravar em relevo páginas inteiras para a impressão de livros e jornais." (Sovik, 2020, p. 3). A autora explica que, em seu sentido secundário, estereótipos são "ideias fixas e simplificadas sobre alguém ou algo, de uma visão chapada" (Sovik, 2020, p. 4). Nesse sentido, os estereótipos são a repetição excessiva de aspectos negativos, atribuindo à população negra meras características redutivas que reforçam os discursos hegemônicos sobre suas identidades. O método de controle por meio da imagem social da pessoa negra está presente em vários produtos midiáticos e tecnológicos. Por meio de narrativas, a branquitude cotidianamente e reiteradamente veicula discursos que desfavorecem alguns grupos de modo a reforçar a dimensão ideológica do racismo e criar cenários que colaboram com a continuidade da opressão.

Assim, o racismo pode ser incorporado às tecnologias, em um espaço de atuação de poder onde, através do significado e por meio de artifícios simbólicos como o discurso e a representação, os grupos dominantes – que não apenas conhecem a linguagem midiática, os artifícios simbólicos pertinentes a essas tecnologias, como têm vantagens históricas da sua herança colonial investindo economicamente na produção de infraestruturas de comunicação e informação – exercem o poder de legitimação social. Mesmo quando supostamente as tecnologias sejam enquadradas como "despolitizadas", "neutras" e "prestativas à sociedade", o racismo se renova e se atualiza, fazendo novas/antigas problemáticas emergirem.

Racismo algorítmico, inteligência artificial e branquitude

Em um panorama onde os dados se configuram como uma nova forma de recurso na dinâmica do capitalismo (Srnicek, 2016), é possível perceber a persistência da clássica lógica de colonização onde o colonialismo de dados (Couldry; Mejias, 2019), através da extração e das relações envolvendo dados, perpetua a lógica de exploração e a violência epistêmica que caracterizam os sistemas coloniais. A principal distinção é que, diferentemente do colonialismo dos séculos anteriores — sustentado por forças como ideologias supremacistas — essa forma de poder se revela através de algoritmos opacos (Couldry; Mejias, 2019). Nesse sentido, as plataformas são conduzidas por grupos que pertencem a estruturas de poder hegemônicos, utilizando dados que evidenciam as desigualdades sociais já existentes (Noble, 2018; O'Neil, 2021).

Reconhecendo que as dinâmicas de poder impactam as tecnologias algorítmicas (Silva, 2022), é essencial destacarmos que o racismo se configura como uma relação de poder que fundamenta desigualdades relacionadas à raça (Almeida, 2019). Nesse sentido, avançamos alinhados com Lima (2022) ao ressaltar que uma das consequências sociais, associada à utilização das tecnologias, é o racismo algorítmico. Silva (2022, p. 1167) descreve o termo como a forma pela qual "a disposição de tecnologias e imaginários sociotécnicos em um mundo moldado pela supremacia branca realiza a ordenação algorítmica racializada de classificação social, recursos e violência em detrimento de grupos minorizados".

Nesse sentido, as experiências on-line são mediadas por processos computacionais que analisam uma ampla massa de dados sobre seus usuários, estruturam como a informação é produzida, acessada, organizada, vista como legítima ou descartada como irrelevante (Ananny, 2016). Embora esses processos computacionais sejam múltiplos, neste trabalho buscamos nos aproximar das questões que envolvem a inteligência artificial (IA) e o uso de *machine learning*.

Pesquisas e opiniões populares sobre essas temáticas têm conquistado consideráveis debates dentro da sociedade científica, das mídias e de conversas do cotidiano. Embora ainda não haja um acordo sobre o que exatamente significa "inteligência artificial", a combinação dessas duas palavras tem sido empregada como um termo abrangente para se referir a máquinas que, se fossem seres humanos, seriam vistas como inteligentes (Mccarthy, 2000). Essa expressão também abrange sistemas computacionais que pensam e agem de maneira similar aos humanos, além de raciocinarem e atuarem de forma racional (Gomes, 2010). Atualmente, diversas IAs são baseadas em *machine learning*. Isso indica que essas aplicações são desenvolvidas com equações estabelecidas previamente, permitindo a organização e execução dos dados conforme necessário (Damaceno, Vasconcelos, 2018). Assim, conseguem elaborar resultados sobre problemas a partir do reconhecimento de padrões presentes em uma base de dados (Oliveira, 2018). Portanto, o êxito dessas aplicações

está intimamente relacionado à qualidade dos dados que foram disponibilizados ou eliminados, permitindo que a máquina processe e forneça os resultados esperados (Costa; Kremer, 2022).

Com base nessas observações surgem reflexões necessárias acerca dos dados que estão sendo utilizados para o processo de aprendizagem de determinadas IAs. Se um sistema de IA entrega um resultado enviesado, como a produção de imagens baseadas em estereótipos racistas ligados à população negra, quais dados foram utilizados para que a máquina entregasse tais desfechos? Por que uma IA entrega imagens recheadas de cenários utópicos destoantes dos registros históricos? Estratégias como a exclusão de categorias sensíveis – como raça, gênero e sexualidade – são propostas como possíveis soluções para os vieses presentes em algoritmos. No entanto, isso não basta, uma vez que há outros fatores que envolvem o racismo e outras formas de preconceito na sociedade (Silva, 2022). Outra alternativa considerada para essa problemática é a transparência algorítmica, mas ela pode não ser eficaz em esclarecer como um algoritmo chegou a variáveis racistas, sexistas ou formas correlatas de opressão. Isso porque, em certos modelos de inteligência artificial, como o das redes neurais artificiais, a forma de operação do algoritmo não permite explicar os procedimentos e passos que levaram a uma determinada decisão (Silveira; Silva, 2020).

Ao examinar o surgimento do racismo em ambientes virtuais, Daniels (2009; 2013) observa que, desde a década de 90, os supremacistas enxergaram a internet como um espaço para suas atividades. Isso se manifestou na criação de sites que disseminavam desinformação sobre líderes históricos da luta pelos direitos civis da população negra e na construção de portais que conectavam internacionalmente grupos extremistas. Assim, concordamos com Noble (2018) ao ressaltar que as perspectivas feministas interseccionais nos estudos digitais podem atuar como uma fundamentação para contestar as narrativas que apresentam a Internet e as plataformas digitais como ambientes democráticos. É fundamental destacar que o racismo nas plataformas, sejam elas de inteligência artificial ou não, não se limita às situações isoladas. Ele se revela de maneira estruturada e intencional, refletindo um sistema de privilégios e poder político, cultural e econômico voltado para certos grupos da sociedade, especialmente a população branca (Tynes *et al.*, 2018).

No cenário em questão é fundamental considerarmos o estudo de Bento (2002), que abordou de maneira significativa as desigualdades sociais entre indivíduos brancos e não brancos a partir do conceito de branquitude. Esse conceito se refere a um ambiente de privilégios raciais, econômicos e políticos no qual a percepção da raça – permeada por significados, experiências e identificações emocionais – desempenha um papel fundamental na organização da sociedade. Nessa perspectiva, como já mencionado, diferentes pesquisadores (Silva, 2020; 2022; Rodrigues *et al.*, 2023; Karam Tietboehl *et al.*, 2024) destacaram a conexão entre o regime da branquitude e

as opressões algorítmicas, como o racismo algorítmico.

A manutenção e replicação dos privilégios associados à branquitude estão intimamente ligadas à dominação colonial e neocolonial, permeando campos que vão da ciência à tecnologia (Silva, 2020). Segundo Bento (2002), a branquitude mantém as hierarquias raciais de modo a estabelecer um acordo entre iguais, especialmente presente nas organizações que têm como principal função a reprodução e conservação dessas estruturas. Nesse sentido, Silva (2020) destaca que organizações como a Ciência e a Tecnologia têm um papel crucial na manutenção e na continuidade dos privilégios relacionados à branquitude. Isso gera uma "dupla opacidade", que se manifesta na forma como os discursos predominantes encobrem tanto os aspectos sociais da tecnologia quanto as discussões sobre a relevância das questões raciais em várias áreas da sociedade, incluindo o setor tecnológico.

Notas metodológicas sobre os casos *Copilot* e *Gemini*

Este artigo adota uma abordagem de estudo de caso múltiplo (Yin, 2009), de natureza descritiva, para analisar dois escândalos envolvendo vieses algorítmicos em sistemas de inteligência artificial desenvolvidos pela Microsoft e Google. A escolha por analisar dois casos, em vez de um único, justifica-se pela busca de uma maior abrangência na identificação de indicativos da branquitude nos produtos tecnológicos. Conforme aponta Yin (2015, p. 28), "apesar de estudos de caso único poderem render insights inestimáveis, a maioria dos estudos de casos múltiplos tem a probabilidade de ser mais forte do que os projetos de caso únicos".

Constata-se uma notória ausência de discussões informacionais relevantes mediadas por parte das empresas responsáveis por tais ferramentas acerca das temáticas em questão, o que, por sua vez, contribui para a opacidade que permeia esse contexto. Portanto, a coleta de dados foi realizada por meio de observação direta (Yin, 2015) em matérias de jornais digitais e publicações em redes sociais virtuais que continham descrições e relatos de entrevistas sobre os escândalos. Os dados coletados consistem em produções textuais e imagens relacionados a cada um dos casos analisados.

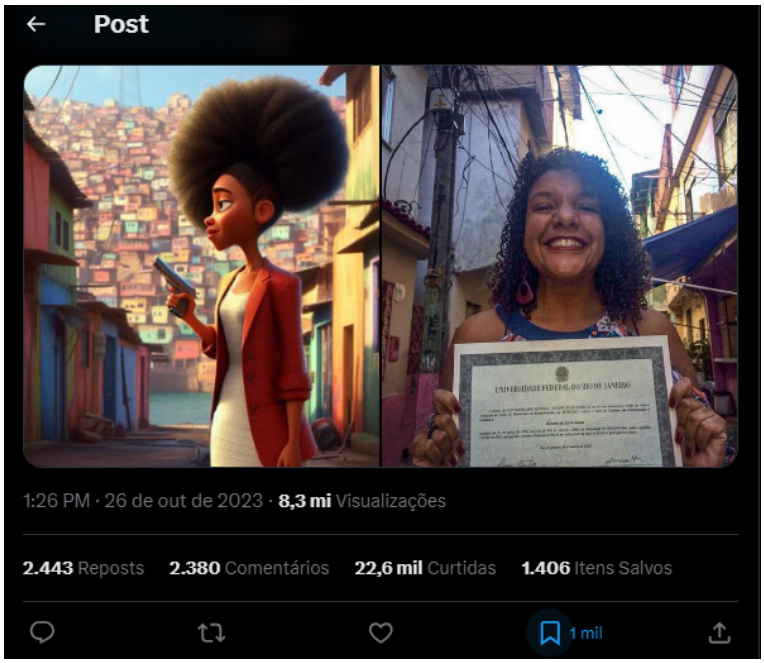
O primeiro caso ocorreu em 2023 com o *Copilot*, de *chatbot* da Microsoft Bing, onde uma deputada brasileira forneceu *prompts* para que a assistente gerasse uma imagem sua, descrita como: uma mulher negra, de cabelos afro, com roupas de estampa africana num cenário de favela. Nesse caso, a assistente entregou a imagem de uma mulher negra associada ao estereótipo de criminalidade. O segundo caso ocorreu em 2024 com o assistente de *chatbot* da Google, o *Gemini*, que também tem a proposta de gerar imagens originais. Nesse caso, um usuário da mídia social X publicou em seu perfil os resultados de *prompts* sobre imagens dos fundadores dos Estados Unidos, dos vikings e do Papa. Ao invés de entregar imagens de homens

brancos, correspondentes aos atores em questão, a IA gerou uma diversidade étnica que não corresponde aos fatos históricos, como, por exemplo, a existência de vikings negros. Os dois casos são apresentados nas Figuras 1, 2, 3 e 4 a seguir.

(a) *Copilot*: o estereótipo de criminalidade

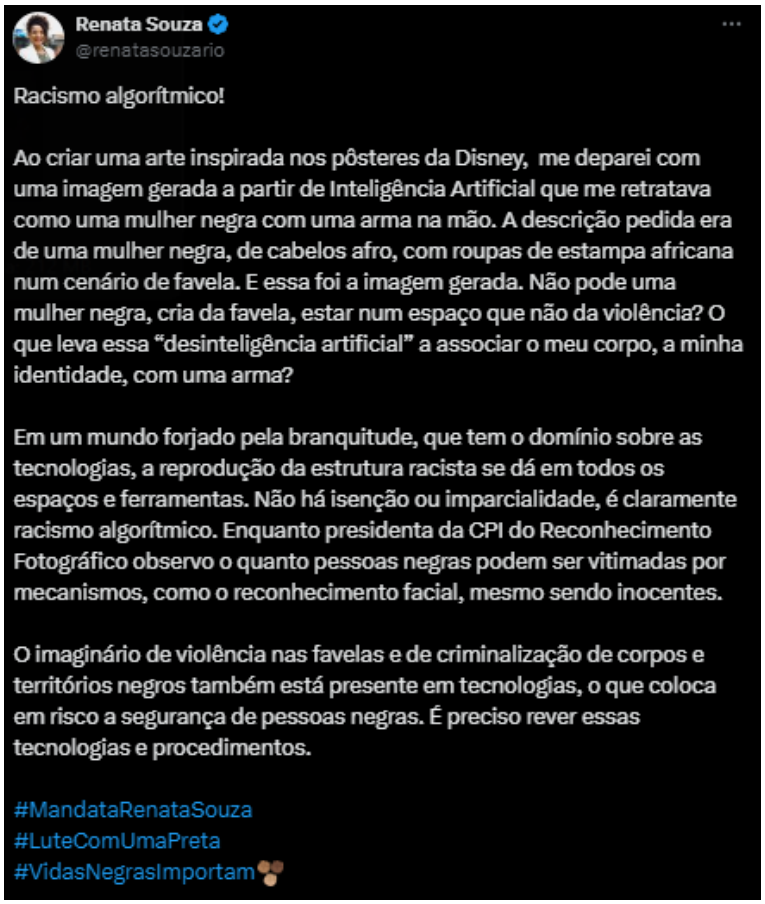
No ano de 2023 algumas empresas de tecnologia já haviam lançado seus assistentes com inteligência artificial (IA), disponíveis em versões gratuitas, capazes de gerar conteúdo em resposta a um *prompt* ou solicitação do usuário. A inteligência artificial generativa é capaz de criar texto, imagem, música e vídeo. Em 26 de outubro de 2023 a deputada estadual Renata Souza Rio, do PSOL (Partido Socialismo e Liberdade) do Rio de Janeiro publicou em uma de suas redes sociais seu relato sobre instruções feitas ao *Copilot*, *chatbot* de IA da Microsoft (Figura 1 e 2), no intuito de gerar uma imagem sua, inspirada nos posters da Disney.

Figura 1 – Post no perfil do X da Renata Souza Rio.



Fonte: Rio (2023).

Figura 2 – Relato da deputada Renata Souza Rio.



Fonte: Rio (2023).

Essa função tem como proposta gerar imagens originais para seus usuários, desse modo é possível fazer releituras, criar uma arte surrealista ou até inspirada em um estilo artístico específico.

Esse foi o pontapé para seu comentário público na rede social X, que trouxe questionamentos sobre o viés algorítmico nas IAs generativas e outros tipos de tecnologias que usam inteligência artificial. Logo o caso repercutiu em notícias, através de sites na internet.

(b) *Gemini*: controvérsias históricas

Em meados do mês de fevereiro de 2024, em meio à corrida desenfreada das empresas de tecnologia por geração de produtos baseados em inteligência artificial, uma considerável controvérsia chamou atenção (Figura 3 e 4). A situação emergiu no debate público quando um usuário da mídia social X publicou em seu perfil resultados gerados através do *chatbot Gemini* decorrentes de *prompts.* Essa postagem foi amplificada por pessoas conhecidas por seus discursos conservadores e posicionamentos contrários à diversidade, como Elon Musk e Jordan Peterson.

Figura 3 – A contradição do soldado alemão nazista, negro.

Fonte: Shamim (2024).

Figura 4 - Post denunciando o viés algorítmico no *Gemini*.

Fonte: Shamim (2024).

Em resposta, a Google – empresa responsável pela IA - anunciou a suspensão temporária da capacidade do *Gemini* de gerar imagens de pessoas. A empresa declarou que as imagens criadas por IA se devem à dedicação da empresa para eliminar preconceitos que, anteriormente, alimentavam estereótipos e comportamentos

discriminatórios. Prabhakar Raghavan, um dos vice-presidentes da empresa, compartilhou uma publicação no blog oficial da Google explicando que a ferramenta *Gemini* estava configurada para contemplar uma variedade de pessoas, porém não estava ajustada para evitar algumas situações inadequadas.

Com a repercussão do caso na mídia, principalmente norte-americana, em um momento em que as discussões sobre representações e viés algorítmico estão cada vez mais em alta, diversos veículos de informação buscaram divulgar detalhes do acontecimento. O jornal The Washington Post entrevistou Margaret Mitchell, ex-co-líder de IA Ética da Google e cientista-chefe de ética da start-up de IA Hugging Face, para discutir o assunto. Mitchell alerta que os resultados em questão poderiam ter sido influenciados por intervenções do Google. Na tentativa acelerada de mitigar o viés, a empresa utilizou o modelo de IA de texto para imagem chamado Imagen 2 e, quando essa capacidade foi incorporada ao *Gemini*, a empresa "ajustou-a". Esses ajustes podem consistir na adição de termos de diversidade étnica aos *prompts* do usuário sem que o mesmo fosse visível na sua tela. Isso significa que os termos anexados podem ser escolhidos aleatoriamente e os *prompts* também podem ter vários termos anexados. Dentro desse contexto, essas correções abordaram o viés com alterações feitas depois que o sistema de IA foi treinado, desconsiderando filtros extremamente necessários – a exemplo das questões históricas – e a curadoria dos dados desde o início.

Estereótipo de criminalidade e controvérsias históricas nos produtos *Copilot* e *Gemini*

No caso do *Copilot* (a), a deputada Renata questiona sobre o que leva a tecnologia a associar o seu corpo e a sua identidade à armas. Resgatando a compreensão sobre identidade abordada anteriormente, entendida por Gomes (2017) como “um fator importante na criação das redes de relações e de referências culturais dos grupos sociais” (Gomes, 2017, p. 41), considera-se que a deputada ofereceu referências suficientes sobre como ela, enquanto mulher negra, gostaria de ser representada pela tecnologia de modo a negar as possibilidades hegemônicas/padronizadas que o *prompt* – com apenas a atribuição “mulher” – poderia gerar.

Ao fornecer essas instruções, Renata indica um traço cultural que se expressa, também, através das suas práticas políticas quando afirma sobre raça e território, seu corpo e sua identidade. Ela faz uso da palavra “favela” para se referir ao território em que se reconhece. Esses são os seus traços de reconhecimento social que a constitui enquanto pessoa e que forma redes de relações Gomes (2017). A deputada realiza o *post* argumentando que: “Em um mundo forjado pela branquitude, que tem o domínio sobre as tecnologias, a reprodução da estrutura racista se dá em todos os espaços e ferramentas”. Essa citação diz sobre como as empresas de tecnologia

estão sob o domínio da branquitude e, por isso, suas ideologias e concepções estarão em todo tipo de produto tecnológico. Como apontado anteriormente, apesar de avanços significativos, a tecnologia existe em circunstâncias sociais, o que não a torna automaticamente acessível, inclusiva, nem tão pouco neutra.

Alinhado com Schwarcz (1988), é persistente a associação entre a pessoa de cor e a noção de violência nas publicações dos jornais impressos. A autora relata que isso se associa de forma tão imediata que a palavra "negro" nos jornais já indicava fatos ruins, incluindo expressões comumente utilizadas que remetem ao negro uma conotação negativa, como "página negra" e "negro crime", que caracterizam acontecimentos violentos (Schwarcz, 1988). Os estereótipos continuam a se repetir nas tecnologias que sucedem, como a inteligência artificial, que promove mais uma vez a representação da pessoa negra como perigosa. Uma mulher negra segurando uma arma cai dentro desse abismo visual de representações sociais baseadas em um imaginário de estigmas. Assim como na imprensa "existe o negro das 'ocorrências policiais', o negro violento que se evadiu, o negro que é centro de notícias escandalosas" (Schwarcz, 1988, p. 95), nas imagens produzidas pelas IAs há o negro estereotipado.

Sobre a criminalização do corpo negro, Renata menciona que "o imaginário de violência nas favelas e de criminalização de corpos e territórios negros também está presente em tecnologias, o que coloca em risco a segurança de pessoas negras. É preciso rever essas tecnologias e procedimentos." O risco de que pessoas negras estejam sempre associadas a estereótipos como esse, diante de tecnologias que prometem segurança pública, nos mostra a importância de se refletir sobre como essa tecnologia será pensada e treinada para executar essa tal segurança. Faz-se necessário a compreensão sobre como os algoritmos são treinados e como as informações usadas em seus treinamentos podem ser uma alternativa para evitar os vieses.

Quanto ao caso do *Gemini* (b), é importante considerar que, em sua maioria, as IAs são ensinadas com base em informações coletadas da internet, predominantemente provenientes dos Estados Unidos e da Europa, o que resulta em uma visão limitada do mundo (Noble, 2018). Assim como os grandes modelos de linguagem funcionam como máquinas de probabilidade ao prever a próxima palavra em uma frase, os geradores de imagens de IA tendem a reproduzir estereótipos na medida em que mostram as imagens frequentemente relacionadas a uma palavra, segundo os usuários americanos e europeus da internet. No entanto, a controvérsia encontrada no *Gemini* vai além dos dados que foram negligenciados durante o processo de aprendizado da IA. Ao considerar a ideia de realizar "ajustes" como uma forma de combater vieses, essas empresas de tecnologia estão optando por métodos menos dispendiosos do que outras intervenções, como: a) filtrar os vastos conjuntos de dados com bilhões de pares de imagens e legendas utilizados no treinamento do modelo; b) o ajuste do modelo no seu ciclo final de desenvolvimento, usando inclusive um feedback humano.

Essa situação nos leva a considerar os estudos de Silva e Cardoso (2017), ao

apontarem que a estrutura e perpetuação da branquitude se dá pela manutenção de privilégios e pela imposição do eurocentrismo como o epicentro da cultura global. É possível identificar esse fenômeno quando observamos a suposta priorização dos dados provenientes do norte global no processo de *machine learning*. Nesse sentido, ao realizar "ajustes" ao invés de um aprofundamento em possíveis soluções e testes realísticos para os vieses, a controvérsia protagonizada pelo *Gemini* coincide com Bastos (2016) ao pontuar que a hegemonia da branquitude não colabora para que os indivíduos brancos passem a questionar seus privilégios bem como se importar com as desvantagens impostas aos demais grupos.

A branquitude pode causar a invisibilidade, a distância e o silenciamento sobre a existência do outro (Bento, 2017). Assim, o ocorrido com o *Gemini* não se distancia desses fenômenos ao emergir em um cenário onde muito se discute sobre a importância da mitigação dos vieses. Nessa mesma linha de raciocínio, portanto, Bento (2017) aponta que a branquitude se expande, se espalha, se ramifica e direciona o olhar do branco; isso fica evidente nesse debate contraditório, onde surge a oportunidade para a intensificação de discussões fundamentadas em ideologias supremacistas e conservadoras a respeito da relevância das formas de representação de minorias em sistemas de inteligência artificial e na sociedade em geral.

Considerações Finais

Este estudo se concentrou na possibilidade de compreender o regime da branquitude em sistemas de inteligência artificial baseados em representações racistas e, ao mesmo tempo, refletir na direção da opressão algorítmica e da não neutralidade das tecnologias digitais, a partir de um olhar crítico-racial que considera as especificidades de dois casos de vieses algorítmicos envolvendo *chatbots* que fazem uso de sistemas de inteligência artificial: o *Copilot* e o *Gemini*.

O escândalo envolvendo o *Copilot* e a deputada Renata Souza Rio mostra a maneira como a *machine learning* se dá ao encontro da branquitude e do imaginário social racista enraizado nas formulações matemáticas – elaboradas por seres humanos – que guiam as decisões automatizadas. A partir da análise da controvérsia envolvendo o *Gemini* é possível dizer que os traços da branquitude e de um imaginário social racista ensejam a busca incessante por uma solução dos vieses algorítmicos através de pequenos ajustes no produto final em detrimento de investimentos em soluções aprofundadas e de maior eficácia, gerando, assim, possibilidades para um fortalecimento de discursos conservadores e contrários à diversidade. Os problemas identificados evidenciam que "a hegemonia da brancura presente em todos os âmbitos sociais não colabora para que os indivíduos brancos passem a questionar seus privilégios bem como se importar com as desvantagens impostas aos demais grupos" (Bastos, 2016, p. 227).

Com base na análise realizada, este estudo identifica um panorama desafiador no que concerne ao combate às opressões algorítmicas. É necessário que as organizações de tecnologia assumam a responsabilidade pelos problemas estruturais inerentes aos seus produtos, reconhecendo-se como agentes partícipes na construção de um cenário social equitativo. Tal reconhecimento implica na elaboração de estratégias para o desenvolvimento de aprendizado de máquina fundamentado em bases de dados diversificadas e treinamentos que incorporem o *feedback* humano.

Destacamos a necessidade de aprimoramentos futuros de um estudo como este, visando às modificações suscetíveis que envolvem os meios, os sujeitos e os dados explorados, o que indica a abertura para estudos posteriores que investiguem as conexões existentes entre o imaginário social racista, a branquitude e a inteligência artificial. Almeja-se que a presente reflexão contribua para o desenvolvimento de investigações futuras.

Artigo submetido em 29/10/2024 e aceito em 04/04/2025.

Referências

ALMEIDA, S. L. **Racismo estrutural** (Feminismos Plurais). São Paulo: Pólen, 2019.

ANANNY, M. Toward an ethics of algorithms: convening, observation, probability, and timeliness. Science, **Technology & Human Values**, v. 41, n. 1, p. 93–117, 2016.

ARAÚJO, E. P. O. Mídia cotidiano e imaginário. *In*: LINS, E. S.; MORAES H. J. P. (Org.). **Informação e Imaginário**. João Pessoa: Editora UFPB, 2019. p.59-69.

BARLOW, J. P. **Declaração de Independência do Ciberespaço**. Dhnet. Davos, Suíça, 8 fev. 1996. Disponível em: https://www.dhnet.org.br/ciber/textos/barlow.htm. Acesso em: 01 jun. 2024.

BASTOS, J. R. B. O lado branco do racismo: a gênese da identidade branca e a branquitude. **Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)**, v. 8, n. 19, p. 211-231, 2016.

BENJAMIN, R. **Race After Technology**: Abolitionist Tools for the New Jim Code. John Wiley & Sons, 2019.

BENTO, M. A. S. **O pacto da branquitude**. São Paulo: Companhia das Letras, 2022.

BENTO, M. A. S. **Pactos narcísicos no racismo**: branquitude e poder nas organizações empresariais e no poder público. Universidade de São Paulo, São Paulo, 2002.

BENTO, M. A. S. Branqueamento e branquitude no Brasil. *In*: CARONE, I.; BENTO, M. A. S. **Psicologia social do racismo**: estudos sobre branquitude e branqueamento no Brasil. Editora Vozes Limitadas, 2017.

BOWKER, G. C.; STAR, S. L. **Sorting things out**: classification and its consequences. Cambridge, MA: MIT Press, 2000.

CASTELLS, M. **A Sociedade em Rede**: Economia, Sociedade e Cultura. v. 1. 6. ed. São Paulo: Paz e Terra, 2013.

COLLINS, E. **New and better ways to create images with Imagen 2**. The Keyword Google. 1 fev. 2024. Disponível em: https://blog.google/technology/ai/google-imagen-2. Acesso em: 20 jun. 2024.

COSTA, R. S.; KREMER, B. 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, 2022. DOI: https://doi.org/10.30899/dfj.v16i1.1316.

COULDRY, N.; MEJIAS, U. **The costs of connection**: how data is colonizing human life and appropriating it for capitalism.1, Stanford: Stanford Press, 2019.

DAMACENO, S. S.; VASCONCELOS, R. O. Inteligência artificial: uma breve abordagem sobre seu conceito real e o conhecimento popular. **Ciências Exatas e Tecnológicas**, v. 5, n. 1, p. 11-16, set. 2018.

DANIELS, J. **Cyber racism**: white supremacy online and the new attack on civil rights. Rowman & Littlefield Publishers, 2009.

DANIELS, J. Race and racism in Internet studies: A review and critique. **New Media & Society**, v. 15, n. 5, p. 695-719, 2013.

DEPUTADA Renata Souza denuncia racismo em plataformas de IA. **Mídia Ninja**. [*s. l.*], 17 nov. 2023. Disponível em: https://midianinja.org/deputada-renata-souza-denuncia-racismo-em-plataformas-de-ia. Acesso em: 11 maio 2024.

GOMES, C. M. D. C. **O que era preto se tornou vermelho**: representação, identidade e autoria negra na imprensa do século XIX por Luiz Gama. Dissertação (Mestrado em Ciências da Comunicação) - Universidade de São Paulo, São Paulo, 2020.

GOMES, D. S. **Inteligência Artificial**: Conceitos e Aplicações. Revista Olhar Científico, [s. l.], v. 1, n. 2, p. 2-5, ago. 2010.

GOMES, N. L. **Alguns termos e conceitos presentes no debate sobre relações raciais no brasil**: uma breve discussão. São Paulo, 13 mar. 2017. Disponível em: https://tinyurl.com/yrez6acu. Acesso em: 10 jul. 2024.

KARAM TIETBOEHL, L.; SZUCHMAN, K. S.; CASAL, C. D.; COSTA, L. A.; DE PAULA, L. R. Especulando narciso: fabulações digitais com a inteligência artificial sobre a branquitude. **Revista Cerrados**, v. 33, n. 64, p. 17–28, 2024.

LIMA, B. D. F. B. **Racismo Algorítmico**: o enviesamento tecnológico e o impacto aos direitos humanos. 2022. 127 f. Dissertação (Mestrado em Direito) – Universidade Federal de Sergipe, São Cristóvão, 2022.

LOPES, F.; LINS, E. S. Mídia cotidiano e imaginário. *In*: LINS, E. S. MORAES H. J. P. (Org.). **Jornalismo**: campo fértil para a compreensão do imaginário social. João Pessoa: Editora UFPB, 2019. p.85 -95.

LORENA, B. M. PIO, I. M. As imagens de controle no contexto brasileiro: os limites e as potencialidades do conceito de Patricia Hill Collins. *In*: Cachel, A. Ferreira C. C. (Org.). **Conexões Humanas**: reflexões do XIII SEPECH-UEL. Londrina: UEL, 2022. p.156-166.

MCCARTHY, J. **What is artificial intelligence?** Stanford, 2000. Disponível em: http://www-formal.stanford.edu/jmc/whatisai.pdf. Acesso em: 20 abr. 2025.

MORAES, M. L. B. Stuart Hall: cultura, identidade e representação. **Revista Educar Mais**, v. 3, n. 2, p. 167-172, 2019.

MOURA, C. **Sociologia do negro brasileiro**. São Paulo: Perspectiva, 2019.

MUSK, E. **Google Gemini is super racist & sexist!** [*s.l.*], 27 fev. 2024. X: @elonmusk. Disponível em: Elon Musk no X: "Google Gemini is super racist & sexist!" / X. Acesso em: 10 jun. 2024.

NEWS FROM GOOGLE. **Pausa na Geração de Imagens pelo Gemini**. 22 fev. 2024. X: @ newsfromgoogle. Disponível em: https://x.com/Google_Comms/status. Acesso em: 5 jun. 2024.

NOBLE, S. U. **Algorithms of oppression**: how search engines reinforce racism. NYU Press, 2018.

NÓBREGA, L. Gemini erra em questões históricas e raciais e Google suspende geração de imagens. **Desinformante**. 23 fev. 2024. Disponível em: https://desinformante.com.br/gemini-erra-google. Acesso em: 3 jul. 2024.

NOVAES, R. R. Imaginário social e educação. *In*: TEVES, N. (Org.). **Um olhar antropológico**. Rio de Janeiro: Gryphus, 1992. p. 122-143.

OLIVEIRA, C. Aprendizado de máquina e modulação do comportamento humano. *In*: SOUZA, J.; AVELINO, R.; SILVEIRA, S. A. **A Sociedade de Controle**: manipulação e modulação nas redes digitais. São Paulo: Hedra, 2018.

O'NEIL, C. **Algoritmos de destruição em massa**. Editora Rua do Sabão, 2021.

RAGHAVAN, P. **Gemini image generation got it wrong. We'll do better**. The Keyword Google. 23 fev. 2024. Disponível em: https://blog.google/products/gemini/gemini-image-generation-issue/. Acesso em: 15 maio 2024.

RESMINI, C. B.; PAGLIARINI, E.; MORAES, J. D. R.; LANDMEIER, L. V. Os desafios sociais na Era da Inteligência Artificial: um enfoque na lacunar equidade racial. **Revista Avant - ISSN 2526-9879**, v. 8, n. Especial, p. 221–238, 2024.

RODRIGUES, J. C.; CHAI, C. G. Inteligência artificial e racismo algoritmo: análise da neutralidade dos algoritmos frente aos episódios de violação de direitos nos meios digitais. **Revista eletrônica [do] Tribunal Regional do Trabalho da 9ª Região**, Curitiba, v. 12, n. 118, p. 92-103, mar. 2023.

SCHWARCZ L. M. **Retrato em branco e negro**: jornais, escravos e cidadãos em São Paulo no final do século XIX. São Paulo: Círculo do Livro, 1988.

SHAMIM, S. **Why Google's AI tool was slammed for showing images of people of colour**. Aljazeera. 9 mar. 2024. Disponível em: https://tinyurl.com/2s3cxefa. Acesso em: 15 jun. 2024.

SILVA, C. M.; CARDOSO, P. J. F. O fim do arco-iris: a branquintude como desafio da luta antirracista no brasil contemporaneo. *In*: MÜLLER, T. M. P.; CARDOSO, L. (org.). **Branquitude**: estudos sobre a identidade branca no brasil. Curitiba: Appris, 2017. p. 243-258.

SILVA, T. 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, 2020.

SILVA, T. **Racismo algorítmico**: inteligência artificial e discriminação nas redes digitais. Editora Sesc SP; 2022.

SILVEIRA, S. A.; SILVA, T. R. Controvérsias sobre danos algorítmicos. **Revista observatório**, v. 6, n. 4, p. 1-17, jul./set. 2020.

SOUZA, R. **Racismo algorítmico!** [...]. [*s.l.*]26 out. 2023. X: @renatasouzario. Disponível em: https://x.com/renatasouzario/status/1717578373826810202. Acesso em: 6 maio 2024.

SOVIK, L. "Através do olhar da representação": sobre o estereótipo e a comunicação. **Heterotopías**, v. 3, n. 6, p. 1-27, 2020. Disponível em: https://tinyurl.com/28am3vds. Acesso em: 20 abr. 2025.

SRNICEK, N. **Platform Capitalism**. Cambridge: Polity Press, 2016.

TEVES, N. Imaginário social e educação. *In*: TEVES, N. (Org.). **O imaginário na configuração da realidade social**. Rio de Janeiro: Gryphus, 1992. p.3-33.

TIKU, N.; SCHAUL, K; CHEN; S. Y. These fake images reveal how AI amplifies our worst stereotypes. **The Washington Post**. [*s.l.*], 01 nov. 2023. Disponível em: AI generated images are biased, showing the world through stereotypes - Washington Post. Acesso em: 3 jul. 2024.

TYNES, B. M. *et al.* From Racial Microaggressions to Hate Crimes: A Model of Online Racism Based on the Lived Experiences of Adolescents of Color. *In*: TORINO, G. C.; RIVERA, D. P.; CAPODILUPO, C. M. Kevin L.; NADAL, K. L.; SUE, D. W. **Microaggression Theory**: Influence and Implications, 2018.

VYNCK, G.; TIKU, N. Google takes down Gemini AI image generator. Here's what you need to know. **The Washington Post**. [*s.l.*], 23 fev. 2024. Disponível em: https://tinyurl.com/4jkxvsy2. Acesso em: 3 jul. 2024.

YIN, R. K. **Case study research, design and methods (applied social research methods)**. Thousand Oaks. California: Sage Publication, 2009.

YIN, R. K. **Estudo de Caso**: Planejamento e métodos. Porto Alegre: Bookman, 2015.