
Entre dados e danos: como a Inteligência Artificial Generativa atualiza as discriminações raciais¹

Michael Felipe Souza Araujo²

Alan César Belo Angeluci³

Universidade de São Paulo - USP

Resumo

O artigo busca desmistificar a ideia de neutralidade das Inteligências Artificiais Generativas (IAG), evidenciando como esses sistemas, ao serem treinados com dados enviesados, reproduzem e atualizam desigualdades estruturais como o racismo. Apesar de seu caráter técnico, a IA não é neutra nem autônoma, mas sim um reflexo de contextos sociais, como a análise do caso da deputada Renata Souza (PSOL) que expõe como a IAG reforça estereótipos ao associar negritude à criminalidade. A falta de diversidade nos dados e no design dos sistemas de IA aprofunda os danos simbólicos e materiais à população negra.

Palavra-chave: comunicação; tecnologia; inteligência artificial; IA generativa; racismo algorítmico.

Introdução

Nos últimos anos a Inteligência Artificial (IA), mais especificamente a Inteligência Artificial Generativa (IAG) angariou milhões de usuários em todo mundo e cada vez mais tem espaço no nosso cotidiano. Além do uso regular nas plataformas de redes sociais, em ferramentas de trabalho e ao incorporar produtos de grandes empresas, a IAG ganhou destaque na mídia por promover discussões em torno dos impactos nas profissões, debates éticos e sobre regulamentação (FERRARI, 2024, p. 84).

Nick Couldry (2024), alerta que as práticas e discursos que adotamos acerca da IA representam uma redefinição da produção do conhecimento e das relações de poder e que esse rearranjo se impõe com a nossa participação cotidiana, mesmo sem que compreendamos as implicações políticas, sociais e econômicas que elas carregam. O sociólogo ainda destaca como o termo embute uma narrativa de marketing para atender interesses de grandes corporações de tecnologia.

¹ Trabalho apresentado no GP Tecnologias e Culturas Digitais, do 25º Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 48º Congresso Brasileiro de Ciências da Comunicação.

² Mestrando do Programa de Pós-graduação em Ciências da Comunicação (PPGCOM) na Escola de Comunicações e Artes (ECA) da Universidade de São Paulo (USP). E-mail: omikeajo@usp.br.

³ Orientador. Professor doutor dos Programas de Pós-Graduação em Ciências da Comunicação (PPGCOM) e Ciência da Informação (PPGCI) da Escola de Comunicações e Artes (ECA) da Universidade de São Paulo (USP). E-mail: aangeluci@usp.br.

Além disso, também são reportadas inúmeras denúncias⁴ de racismo, sexismo, entre outras injustiças sociais relacionadas ao uso das IAGs (VIEIRA, 2025). Assim, casos de discriminação e reprodução de estereótipos raciais em decorrência do uso de IA, além ganharem visibilidade na mídia e nas redes sociais, também funcionam como gatilhos para novas formas de violência online contra pessoas negras e grupos marginalizados.

Esses episódios não podem ser compreendidos isoladamente, pois evidenciam como as IAGs operam a partir de dados enviesados e de contextos históricos marcados por desigualdades. Logo, este trabalho adota uma abordagem qualitativa e exploratória, com base em revisão bibliográfica recente que discute IA e questões raciais e também na análise do caso inédito da deputada Renata Souza (PSOL-RJ), cuja repercussão midiática e nas redes sociais permite ilustrar como a IAG pode atualizar e reforçar estereótipos raciais sob a aparência de neutralidade tecnológica.

Desmistificando a neutralidade da IA

A OECD⁵, em sua plataforma *OECD.IA Policy Observatory*⁶ dedicada à promoção de uma Inteligência Artificial mais confiável e centrada no humano, define:

Um sistema de IA é um sistema baseado em máquina que, para objetivos explícitos ou implícitos, infere, a partir das informações recebidas, como gerar resultados, como previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos ou virtuais. Diferentes sistemas de IA variam em seus níveis de autonomia e adaptabilidade após a implantação.

A plataforma ainda dispõe de um diagrama (Figura 1), que ilustra o funcionamento desses sistemas, considerando o ambiente físico e virtual como contexto.



⁴Acompanhe o mapeamento de danos e discriminação algorítmica, do projeto Desvelar. Disponível em: <https://desvelar.org/casos-de-discriminacao-algoritmica/>. Acesso em: 7 jun. de 2025.

⁵ Sigla em inglês para Organização para a Cooperação e Desenvolvimento Económico.

⁶ Disponível em: <https://oecd.ai/en/>.

Figura 1: Funcionamento do sistema de IA. Fonte: OECD.AI Policy Observatory.

Carla Vieira (2025, p. 19) especifica que os sistemas de IA Generativa são construídas a partir das Redes Neurais Adversárias (GANs, da sigla em inglês), cujo objetivo é a aprendizagem e a geração autônoma de diferentes formatos como imagens, vozes, textos, vídeos, áudios, etc. a partir de grandes volumes de dados. A autora ainda acrescenta que as redes neurais GANs “conseguem aprender através de identificação de padrões e agrupamentos dentro do conjunto de dados” (VIEIRA, 2025, p. 20).

Mesmo com todo caráter técnico e lógico empregado por essas definições, o contexto é mencionado não apenas como parte do processo ou do ambiente influenciado pelos *outputs* gerados, a partir dos dados de treinamento; mas também como agente influenciador da base de dados que serve como *input* para o sistema de IA:

Se é fato que dados de treinamento enviesados produzem erros de precisão do sistema, leituras erradas ou incompletas, também existe um segundo fenômeno na reprodução do racismo, que é quando a tecnologia produz interpretações da realidade que expressam os valores discriminatórios embutidos na sociedade. Ou seja: as respostas da máquina são expressões lógicas, ainda que injustas, do aprendizado classista, racista e sexista da máquina. (MONAGREDA, 2024, p. 144).

Essa constatação tem levado pesquisadores a recusar o uso acrítico do termo “Inteligência Artificial”, já que conceito “pode tanto subestimar a complexidade da inteligência humana quanto apagar todas as camadas de apropriação de trabalho incorporada em sistemas algorítmicos” (SILVA, 2023, n.p.) e Couldry (2024) ainda destaca como a IA é “apenas probabilística, não é criativa.”.

Dessa forma, uma ideia de imparcialidade que é empregado ao discurso tecnológico é ilusória. Os dados não são isentos de valores sociais, culturais e políticos (SILVA, 2022, n.p.), pois são operadas dentro de contextos sociais que refletem desigualdades estruturais. Na obra “Racismo algorítmico: inteligência artificial e discriminação nas redes digitais”⁷, Tarcízio Silva (2022), afirma que a ideia de neutralidade da IA serve para ocultar relações de poder e perpetuar discriminações algorítmicas:

A compreensão de muitas manifestações do racismo [...] é especialmente difícil quando se trata de tecnologias algorítmicas de comunicação, que trazem

⁷ Obra com acesso aberto em: <https://racismo-algoritmico.pubpub.org>.

novas camadas de opacidade em seu funcionamento. Se a tecnologia é erroneamente enquadrada e percebida como neutra, a tal equívoco se soma a negação do racismo como fundante de relações e hierarquias sociais em países como o Brasil.

A crítica a este ideal de neutralidade, também fundamenta-se no conceito de "dupla opacidade", que descreve a negação simultânea do racismo e da política no desenvolvimento tecnológico (SILVA, 2022, n.p.), ignorando os valores e interesses que influenciam o desenvolvimento, o treinamento e a aplicação desses sistemas, criando uma camada adicional de invisibilidade sobre as desigualdades e violências à população negra, não somente nos ambientes digitais, por meio do racismo online, por exemplo, mas também na realidade, como o reconhecimento facial (BARROS, 2025, p. 16) usado para o encarceramento de homens negros.

Vieses de raça, danos e riscos da IA Generativa

A discussão sobre os danos e riscos da IAG é fundamental, e a opacidade presente nos sistemas de IA, que disfarça vieses sob uma falsa neutralidade, revela a urgência de analisar como essas tecnologias são desenvolvidas e consumidas. Não obstante, tais problemáticas perpassam todas as etapas da tecnologia, seja na camada física, lógica ou de conteúdo (RICAURTE QUIJANO, 2023, s.p.).

Nesse sentido, buscar uma tipologia dos danos e riscos da IAG, a partir dessas camadas, torna-se cada vez mais importante e urgente. Rodrigues (2025, p. 47) aponta que esses danos e riscos se manifestam “a partir daqueles associados aos (I) dados que são utilizados em IAs generativas, ao (II) trabalho humano envolvido, aos (III) impactos ambientais decorrentes e pelos seus (IV) resultados e dados gerados”.

Os vieses discriminatórios são inseridos por meio dos dados de treinamentos, que são frequentemente coletados da internet. Se esses dados contêm representações estereotipadas ou desiguais de grupos racializados, os modelos de IA aprendem e reproduzem esses mesmos padrões, que incorporam estereótipos e preconceitos já existentes: “como o treinamento inclui textos e imagens que retratam pessoas negras em papéis estereotipados, os modelos continuam reforçando essas representações nos conteúdos gerados.” (VIEIRA, 2025, p. 22).

Como já visto anteriormente, a aprendizagem da IAG não é neutra, seja pelo fato que pode estar contaminado pelos hábitos e vieses dos usuários, que também alimentam os sistemas, além da falta de diversidade nas equipes de desenvolvimento, predominantemente homens do Norte Global, com ideologias ligadas a lógica colonial, racista, classista e machista (FIGUEIREDO; ALECRIM NETO, 2024, p. 7). A falta de diversidade nos dados, mas também no design desses sistemas reflete nas “formas de opressão e exclusão que condicionaram a constituição tanto dos datasets quanto do software, incluindo aí a pouca diversidade das equipes de desenvolvimento.” (SILVA, 2022, n.p.).

Diante disso, os chamados *outputs*, ou seja, os resultados gerados pela IAG, podem reforçar estereótipos, disseminar desinformação e perpetuar a discriminação. Figueiredo e Alecrim Neto (2024), descrevem um estudo realizado com o Leonardo.AI⁸, demonstra a limitação da IAG em produzir imagens não estereotipadas de pessoas negras, as retratando em contextos de sofrimento ou pobreza, enquanto crianças brancas eram retratadas sorrindo e brincando (FIGUEIREDO; ALECRIM NETO, 2024, p. 8).

Entretanto, a automatização das opressões transforma a violência simbólica da representatividade negativa e estereótipos em danos concretos que afetam a cidadania e a vida da população negra. Logo, a IA funciona como uma atualização de alta tecnologia do processo repressivo do Estado, potencializando a vigilância e a interdição física sobre pessoas negras e suas comunidades (MONTEIRO, 2025, p. 62). Monagreda (2024, p. 112) afirma que os sistemas de reconhecimento facial, usados em segurança pública, têm altas taxas de erro contra pessoas negras e migrantes, levando a "falsos positivos".

Dessa forma, em uma sociedade, baseada num padrão de branquitude, patriarcado e colonialidade, resulta em uma tecnologia que espelha as estruturas de desigualdades históricas e sociais da realidade (FIGUEIREDO; ALECRIM NETO, 2024, p. 6). Portanto, os vieses de raça presentes nas bases de dados, que alimentam as IAGs atualizam as desigualdades estruturais, que resultam na discriminação de grupos racializados, seja por invisibilização, hipervisibilidade ou estereotipação, perpetuando o

⁸ Leonardo.AI é uma plataforma de IAG que permite a criação de imagens, vídeos e outros conteúdos visuais a partir de comandos de texto. Disponível em: <https://leonardo.ai/>.

racismo estrutural de forma automatizada e opaca (SILVA, 2022, n.p.), criando sistemas que mantêm exclusões e opressões, dentro e fora do ambiente digital.

Violências simbólicas e imagens de controle: o caso da deputada Renata Souza

Suzana Varjão (2008) aponta que a falta de representatividade, juntamente com o reforço de estereótipos nos meios de comunicação e em produtos culturais, não implica apenas no reconhecimento estético, mas também na materialidade da violência simbólica e material contra a população negra:

Hierarquia [racial] que aprisiona o negro nas tramas de um processo que lhe é, a todos os títulos, desfavorável, com prejuízos permanentes a uma prática de cidadania. E tramas com engrenagens nem sempre visíveis. (Varjão, 2008, p. 168).

A autora aponta o cenário onde destaca a relação entre o racismo e a performance dessa violência na mídia. Com as novas tecnologias, e consequentemente a IA, fazendo cada vez mais parte do nosso cotidiano, os meios de comunicação também se atualizaram, desde a popularização da internet nos anos 2000, assim como das redes sociais na década de 2010. Por consequência, com a plataformização da vida e, principalmente, dos meios de comunicação e dos produtos culturais, também resvalam nos tempos atuais, com a popularização da Inteligência Artificial:

a penetração de infraestruturas, processos econômicos e estruturas governamentais de plataformas em diferentes setores econômicos e esferas da vida. E, a partir da tradição dos estudos culturais, concebemos esse processo como a reorganização de práticas e imaginações culturais em torno de plataformas (POELL, NIEBORG e VAN DIJCK, 2020, p. 5).

O caso da deputada estadual negra do Rio de Janeiro, Renata Souza (PSOL), ocorrido em outubro de 2023, provocou grande repercussão na mídia por envolver a IA e vieses raciais. Ao participar da *trend*⁹ “Disney Pixar”, usando a ferramenta de IAG da Microsoft, ela obteve a ilustração de uma mulher negra com cabelo afro segurando uma arma em um cenário de favela (Figura 2), sem que houvesse a palavra “arma” no *prompt*¹⁰ solicitado. Em publicação feita no X (antigo Twitter), Renata Souza (2023) denunciou o caso de racismo algorítmico e questionou: “Não pode uma mulher negra,

⁹ Representa um tipo de conteúdo ou estilo que ganha popularidade rapidamente, sendo reproduzido e adaptado por muitos usuários. Fonte: Portal Trend. Disponível em: <https://portaltrend.com.br/o-que-e-trend/>.

¹⁰ É um texto em linguagem natural que solicita que a IA generativa execute uma tarefa específica. Fonte: AWS.

cria da favela, estar num espaço que não da violência? O que leva essa ‘desinteligência artificial’ a associar o meu corpo, a minha identidade, com uma arma?’”.



Figura 2: Imagem gerada pela IA da Microsoft (esq.) e foto da deputada Renata Souza segurando seu diploma (dir.). Fonte: Perfil da deputada no X.

Esse caso evidencia a dinâmica que vem sendo apontada no trabalho, incluindo como a IAG, ao associar a negritude e a favela com a criminalidade, representada pela arma, ressalta e contribui para as imagens de controle, conceito de Patricia Hill Collins (2019), que reforça como estereótipos, principalmente de mulheres negras, culminam em violência simbólica:

Trata-se de uma violência simbólica e discursiva estabelecida pelos opressores que balanceiam raça, classe, gênero, sexualidade e agressividade buscando delimitar um local social pejorativo para a mulher negra. As imagens de controle articulam-se com símbolos no imaginário. Significa dizer que, no inconsciente, determinadas expectativas de como os corpos devem performar discursiva e visualmente e analisar quais os seus papéis. (KLEIN; COSTA, 2023, p. 192)

Entretanto, no caso de Renata Souza, ainda se destacam outras camadas da violência online. Após o relato no X, a deputada foi contestada e desacreditada por outros usuários, porque ao reproduzirem o *prompt* utilizado a imagem gerada não apresentava uma arma, ao menos que solicitasse (Figura 3). Esse desencadeamento de discriminações e violências ainda têm base nas imagens de controle, pois “são construídas para fazer com que o racismo, o sexismo, a pobreza e outras formas de injustiça social pareçam tão normais quanto inevitáveis.” (BARROS, 2025, p. 11).

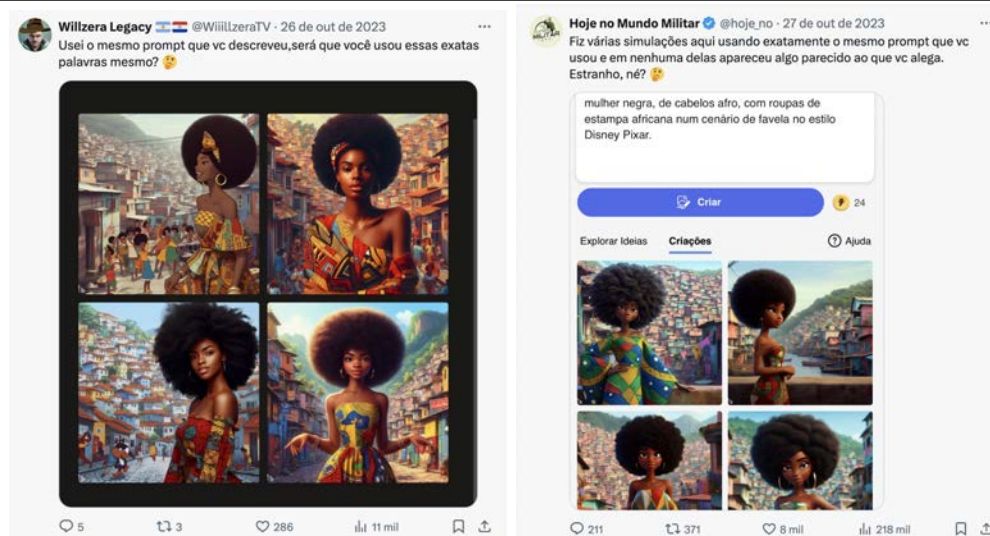


Figura 3: Respostas de usuários contestando a denúncia da deputada. Fonte: Perfil da deputada no X.

Uma outra problemática derivada do caso Renata Souza, perpassa sobre as diretrizes da plataforma, que deveria ser responsável por fazer checagens e auditorias dos conteúdos publicados, e não depositar nas mãos dos usuários tal ação. Isso porque, foi adicionado à publicação da deputada, uma “nota de comunidade”, recurso onde os próprios usuários da rede avaliam e contextualizam o teor das publicações, possibilitando a inserção de uma contextualização enviesada, por parte dos usuários, incluindo uso político e propagação de desinformação.

Considerações finais

A análise desenvolvida neste trabalho demonstra que os sistemas de IA são alimentados por dados enviesados e desenhados por equipes não diversas. A IAG reforça estereótipos e legitima práticas de discriminação, com impactos simbólicos e materiais. A ideia de neutralidade da tecnologia (SILVA, 2022), ajuda a compreender como essas violências permanecem encobertas, dificultando tanto sua identificação quanto a formulação de respostas eficazes.

Nesse sentido, as IAG não apenas reproduzem, mas intensificam a performance de desigualdades, seja em peças visuais ou decisões automatizadas. Ao associar corpos racializados a contextos de violência ou subalternidade, a IAG se torna um vetor das imagens de controle (COLLINS, 2019; KLEIN; COSTA, 2023), agora reformuladas por lógicas algorítmicas. Esses efeitos operam não apenas na esfera simbólica, mas também

na política, que resulta em impactos materiais na segurança e na cidadania das pessoas negras e outros grupos marginalizados.

Diante disso, torna-se urgente a construção de caminhos para mitigar os danos provocados por esses sistemas. Diante deste cenário, a solução não reside na rejeição total da tecnologia, mas em um esforço contínuo de descolonização da IA e de outras tecnologias digitais (AGUIAR; SILVA, 2024, p. 115). Em seu trabalho, Aguiar e Silva (2024), realizam um mapeamento de grupos historicamente marginalizados que propõem ressignificações para a IA a partir de experiências sociotécnicas decoloniais.

Por isso a importância de iniciativas como o projeto "Desvelar"¹¹, que propõe como objetivo¹² principal “contribuir para os desenvolvimentos e debates sobre tecnologias digitais, em especial políticas públicas e regulamentação relacionados à IA no Brasil”, com ética, transparência e promoção dos direitos humanos, entretanto também sob a ótica racial (SILVA; SILVA, 2023). É preciso enfrentar a atualização do racismo com uma agenda antirracista e decolonial, que reprograma as tecnologias em favor da justiça social.

Referências

- AGUIAR, Carlos Eduardo Souza; SILVA, Dayana K. Melo da. Tecnologia e decolonialidade: arranjos insurgentes e a questão das cosmotécnicas. *Tecnologia e Sociedade*, Curitiba, v. 20, n. 62, p. 114–127, out./dez. 2024. Disponível em: <https://bit.ly/3TsP4w0>. Acesso em: 18 jun. 2025.
- BARROS, Thiane Neves. A influência das imagens de controle na I.A. generativa. In: SILVA, Tarcízio (org.). **Inteligência artificial generativa: discriminação e impactos sociais**. 2. ed. São Paulo: LiteraRUA, 2025. p. 10–18.
- COLLINS, Patricia Hill. **Pensamento Feminista Negro: conhecimento, consciência e a política do empoderamento**. Trad. de Jamille Pinheiro Dias. São Paulo: Boitempo, 2019.
- COULDRY, Nick. IA como mito e colonialismo de dados. In: Reunião Magna da ABC 2024. Rio de Janeiro: Academia Brasileira de Ciências, 2024. Disponível em: <https://youtu.be/zNhuyFinjGk>. Acesso em: 13 jun. 2025.
- FERRARI, Pollyana. Desinformação, IAGs e colonialismo digital. In: FERRARI, Pollyana (org.). **Descendentes de Eliza: os impactos da inteligência artificial generativa no mercado de trabalho, na desinformação, nas artes e no pensamento crítico**. Cachoeirinha: Fi, 2024. p. 84–95. Disponível em: <https://bit.ly/40KGcFZ>. Acesso em: 10 jun. 2025.

¹¹ Projeto Desvelar - justiça racial na inteligência artificial e TICs. Disponível em: <https://desvelar.org/>.

¹² Sobre. Disponível em: <https://desvelar.org/sobre/>. Acesso em: 19 jun. de 2025.

FIGUEIREDO, Carolina; ALECRIM NETO, Ivan da Costa. Aplicação da inteligência generativa na produção de imagens de sujeitos escravizados: um experimento-piloto de promptografia. **Triade: Comunicação, Cultura e Mídia**, Sorocaba, v. 12, n. 25, 2024. Disponível em: <https://bit.ly/44QVNWZ>. Acesso em: 14 jun. 2025.

KLEIN, Ana Maria; COSTA, Juliana dos Santos. Interseccionalidade e imagens de controle: os conceitos de raça, gênero e infância e a constituição da autoidentificação de meninas negras. **Revista Educação e Emancipação**, São Luís, v. 16, n. 3, p. 174–202, set./dez. 2023. Disponível em: <https://bit.ly/46tUcaP>. Acesso em: 18 jun. 2025.

MONAGREDA, Johanna K. Por que falar de raça quando falamos de dados pessoais, inteligência artificial e algoritmos? In: GONÇALVES, Adriana; TORRE, Luísa; MELO, Paulo Victor (eds.). **Inteligência artificial e algoritmos: desafios e oportunidades para os media**. Covilhã: LabCom, 2024. p. 103–134. Disponível em: <https://bit.ly/4kq5Jve>. Acesso em: 15 jun. 2025.

MONTEIRO, Pedro Diogo Carvalho. Pensando a I.A. generativa na arquitetura racial-punitiva do Estado. In: SILVA, Tarcízio (org.). **Inteligência artificial generativa: discriminação e impactos sociais**. 2. ed. São Paulo: LiteraRUA, 2025. p. 59–66.

POELL, Thomas; NIEBORG, David; VAN DIJCK, José. **Plataformização. Fronteiras - estudos midiáticos**, São Leopoldo, v. 22, n. 1, p. 2–10, jan./abr. 2020. Disponível em: <https://bit.ly/46CNcIG>. Acesso em: 18 jun. 2025.

RICAUARTE QUIJANO, Paola. Resistencia como reexistencia: la defensa del cuerpo-territorio en la sociedad algorítmica. **Pléyade (Santiago)**, Santiago, n. 32, dez. 2023. Disponível em: <https://bit.ly/44Dy2AD>. Acesso em: 16 jun. 2025.

RODRIGUES, Fernanda. Riscos e danos associados à I.A. generativa e a busca por uma tipologia. In: SILVA, Tarcízio (org.). **Inteligência artificial generativa: discriminação e impactos sociais**. 2. ed. São Paulo: LiteraRUA, 2025. p. 47–58.

SILVA, Tarcízio. **Racismo algorítmico: inteligência artificial e discriminação nas redes digitais**. PubPub, 2022. Disponível em: <https://bit.ly/460dzIy>. Acesso em: 8 jun. 2025.

SILVA, Tarcízio; SILVA, Fernanda dos Santos Rodrigues (orgs.). Lentes antirracistas sobre regulação de inteligência artificial. [S. l.]: Desvelar, 2023. Disponível em: <https://bit.ly/4ljoJgk>. Acesso em: 18 jun. 2025.

SILVA, Tarcízio. Como circular informações sobre IA, riscos e governança. 2024. Disponível em: <https://bit.ly/4lGtoZl>. Acesso em: 8 jun. 2025.

SOUZA, Renata. Racismo algorítmico! Ao criar uma arte inspirada nos pôsteres da Disney [...]. [S. l.], 26 out. 2025. X: @renatasouzario. Disponível em: <https://bit.ly/4lgwjrm>. Acesso em: 18 jun. 2025.

VARJÃO, Suzana. **Micropoderes, macroviolências**. Salvador: EDUFBA, 2008. Disponível em: <https://bit.ly/407ZTHL>. Acesso em: 17 jun. 2025.

VIEIRA, Carla. Como os vieses entram na I.A. generativa? In: SILVA, Tarcízio (org.). **Inteligência artificial generativa: discriminação e impactos sociais**. 2. ed. São Paulo: LiteraRUA, 2025. p. 19–26.