

A Lógica do Esquecimento: poder, governança de dados e a crítica decolonial da inteligência artificial

The Logic of Forgetting: power, data governance, and the decolonial critique of artificial intelligence

Jefferson Igor Duarte Silva  

jefferson.igorbr@gmail.com

Instituto Federal do Rio Grande do Norte, Natal, RN, Brasil

Akynara Aglaé Rodrigues Santos da Silva Burlamaqui  

akynara@gmail.com

Universidade Federal Rural do Semi-Árido, Mossoró, RN, Brazil

Aquiles Medeiros Filgueira Burlamaqui  

aquilesburlamaqui@gmail.com

Universidade Federal Rural do Semi-Árido, Mossoró, RN, Brazil

Resumo

A expansão da Inteligência Artificial (IA) em domínios críticos como saúde e segurança levanta questões urgentes sobre equidade e justiça. Este artigo argumenta que vieses algorítmicos não são meros erros técnicos, mas manifestações de desigualdades estruturais históricas, refletidas em bases de dados que funcionam como artefatos sociotécnicos. A pesquisa investiga dois mecanismos centrais de exclusão: a representação distorcida, que inclui grupos de modo estereotipado, e a invisibilidade seletiva, que apaga sistematicamente populações inteiras dos registros. Através de um quadro teórico que articula os Estudos Críticos de Algoritmos, o Feminismo de Dados e a crítica Decolonial, analisamos um mosaico de casos emblemáticos, incluindo reconhecimento facial, recrutamento automatizado e a omissão de dados LGBTQIA+. Demonstra-se a existência de um ciclo de retroalimentação da desigualdade, argumentando que soluções técnicas de "debiasing" são insuficientes. Por fim, o artigo propõe diretrizes estruturais para a justiça algorítmica, focadas na soberania dos dados, na contestabilidade efetiva e na governança democrática.

Palavras-chave

Inteligência Artificial. Vieses Algorítmicos. Invisibilidade Social. Justiça Algorítmica.

Abstract

The expansion of Artificial Intelligence (AI) into critical domains like healthcare and security raises urgent concerns about equity and justice. This article argues that algorithmic biases are not mere technical errors but manifestations of historical structural inequalities, reflected in datasets that function as sociotechnical artifacts. The research investigates two central mechanisms of exclusion: distorted representation, which includes groups in a stereotypical manner, and selective invisibility, which systematically erases entire populations from the records. Through a theoretical framework that articulates Critical Studies of Algorithms, Data Feminism, and the Decolonial critique, we analyze a mosaic of emblematic cases, including facial recognition, automated recruitment, and the omission of LGBTQIA+ data. It is demonstrated the existence of a feedback loop of inequality, arguing that technical solutions of "debiasing" are insufficient. Finally, the article proposes structural guidelines for algorithmic justice, focused on data sovereignty, effective contestability, and democratic governance.

Linguagem em Foco

Revista do Programa de Pós-Graduação em Linguística Aplicada da UECE

FLUXO DA SUBMISSÃO

Submissão do trabalho: 28/07/2025

Aprovação do trabalho: 23/10/2025

Publicação do trabalho: 09/12/2025



10.46230/lef.v17i3.16015

COMO CITAR

SILVA, Jefferson Igor Duarte *et al.* A Lógica do Esquecimento: poder, governança de dados e a crítica decolonial da inteligência artificial. **Revista Linguagem em Foco**, v.17, n.3, 2025. p. 70-90. Disponível em: <https://revistas.uece.br/index.php/linguagememfoco/article/view/16015>.

Distribuído sob



Verificado com

Plagius
Detector de Plágio

tion, which includes groups in stereotyped ways, and selective invisibility, which systematically erases entire populations from records. Using a theoretical framework that articulates Critical Algorithm Studies, Data Feminism, and the Decolonial critique, we analyze an analytical mosaic of emblematic cases, including facial recognition, automated recruitment, and the omission of LGBTQIA+ data. We demonstrate the existence of a feedback loop of inequality, arguing that purely technical 'debiasing' solutions are insufficient, often masking the political choices embedded within them. Finally, the paper proposes structural guidelines for algorithmic justice, focusing on data sovereignty, effective contestability, and democratic governance.

Keywords

Artificial Intelligence. Algorithmic Bias. Social Invisibility. Algorithmic Justice.

Introdução

A Inteligência Artificial (IA) tornou-se uma força onipresente que molda decisões em múltiplos setores da vida contemporânea. De triagens de currículos a diagnósticos médicos, seus algoritmos operam com autonomia sem precedentes. Contudo, sua aparente objetividade oculta a dependência dos dados que os treinam — cuja qualidade, diversidade e equidade são determinantes para a justiça e eficácia das aplicações (Gupta, 2021).

Parte-se do pressuposto de que bases de dados são construções sociais que refletem relações de poder, valores e preconceitos (Orr; Crawford, 2024). A problemática central é dupla: a representação distorcida, em que grupos marginalizados são incluídos de modo estereotipado, e a invisibilidade seletiva, que torna certas populações indetectáveis para os sistemas de IA (Carvalho; Barreto, 2021).

A relevância desta pesquisa reside na urgência de enfrentar as consequências desses fenômenos. Vieses de gênero, raça, classe e outros, quando embutidos em algoritmos, ampliam desigualdades e criam novas barreiras de acesso (De Lima Viana; Sperandeo De Macedo, 2024). A invisibilidade, por sua vez, nega a existência e as necessidades de certos grupos, deixando-os fora do alcance de serviços e políticas públicas baseadas em IA (Wang; Ramaswamy; Russakovsky, 2022).

Neste contexto, o objetivo é analisar criticamente as origens e os impactos das representações e invisibilidades nas bases de dados de IA, articulando os Estudos Críticos de Algoritmos, o Feminismo de Dados e as perspectivas pós-coloniais sobre tecnologia.

1 Fundamentação teórica

Este artigo contrapõe-se à concepção tecnicista do viés algorítmico, que

o reduz a um erro estatístico corrigível (Barocas; Hardt; Narayanan, 2019). Em seu lugar, propõe-se um quadro analítico crítico que entende a Inteligência Artificial como um artefato sociotécnico, cujas operações e efeitos são inseparáveis das estruturas de poder e das condições históricas que os produzem (Crawford, 2021; Pasquale, 2015). Nessa perspectiva, a tecnologia não é neutra, mas resultado de escolhas políticas, econômicas e epistemológicas.

A investigação articula três vertentes teóricas — governança algorítmica, feminismo de dados e perspectiva pós-colonial — que, em conjunto, permitem uma análise multinível do fenômeno, integrando dimensões técnicas, políticas e simbólicas.

1.1 A governança algorítmica sob a ótica dos estudos críticos

A discriminação algorítmica raramente opera por um único eixo. Raça, gênero, A base desta análise repousa nos Estudos Críticos de Algoritmos, que buscam “abrir a caixa-preta” da governança algorítmica (Pasquale, 2015). Essa abordagem desmonta a ilusão de objetividade computacional, expondo o código como sistema de poder que classifica e disciplina a realidade social. Os algoritmos não apenas refletem o mundo — eles o performam, automatizando decisões que produzem novas e opacas formas de discriminação.

Ruha Benjamin (2019) descreve esse fenômeno como o “novo código Jim Crow”, aludindo às leis de segregação racial, enquanto Noble (2018) identifica “algoritmos de opressão” que perpetuam racismo e sexismo sistêmicos. Assim, compreender o algoritmo implica compreender seus efeitos políticos e epistemológicos.

1.2 O ciclo de vida dos dados sob a lente do feminismo de dados

Uma crítica centrada apenas no algoritmo é limitada; é necessário examinar sua matéria-prima: os dados. O Feminismo de Dados propõe politizar todo o ciclo de vida da informação, questionando quem coleta, classifica e interpreta (D’ignázio; Klein, 2020).

Sob uma lente interseccional, revela-se que as decisões sobre o que e quem contar são atos políticos que tornam certos corpos visíveis e apagam outros. Essa abordagem amplia a análise da representação distorcida para incluir a invisibilidade como mecanismo de poder, evidenciando que a ausência de dados também é forma de exclusão.

1.3 A geopolítica da IA: colonialismo e epistemologias pós-coloniais

A perspectiva pós-colonial é essencial para deslocar o foco do Norte Global e evidenciar que a racionalidade algorítmica, tida como universal, é produto de epistemologias ocidentais específicas (Roche; Wall; Lewis, 2023). Essa lente situa a IA nas relações geopolíticas de produção de conhecimento, denunciando o colonialismo e extrativismo de dados (Couldry; Mejias, 2019), em que modelos desenvolvidos em centros hegemônicos são impostos a contextos periféricos.

Ao articular essas três correntes, o artigo propõe um modelo teórico integrado capaz de compreender a desigualdade algorítmica não como falha técnica, mas como consequência lógica de uma ecologia sociotécnica globalmente hierarquizada.

2 Metodologia: Justificativa Teórica e Mosaico Analítico dos Casos

A pesquisa adota uma abordagem qualitativa e interpretativista, utilizando o estudo de caso como principal ferramenta para revelar as múltiplas camadas do fenômeno investigado. A seleção dos casos foi deliberada, buscando compor um mosaico analítico no qual cada caso ilumina uma dimensão distinta da desigualdade algorítmica, indo além da análise de falhas técnicas isoladas.

A escolha metodológica está alinhada aos referenciais teóricos do trabalho. Sob a lente dos Estudos Críticos de Algoritmos, cada caso é interpretado como um sistema de poder e governança a ser desconstruído. O Feminismo de Dados orienta a análise das dinâmicas de exclusão e categorização — questionando quem é contado, quem é ignorado e quem define as categorias. Já a perspectiva pós-colonial permite compreender como a importação acrítica de tecnologias e o extrativismo de dados reproduzem dependências no contexto brasileiro e global.

Essas lentes teóricas fundamentam três estudos de caso: (1) viés interseccional em reconhecimento facial, (2) viés de gênero em algoritmos de recrutamento e modelos de linguagem, e (3) racismo algorítmico e vigilância digital na segurança pública brasileira. O conjunto busca articular teoria e empiria, compreendendo a desigualdade algorítmica como fenômeno técnico, político e epistêmico.

2.1 Estrutura do Mosaico Analítico

O mosaico analítico oferece uma visão multifacetada.

O primeiro caso revela como raça e gênero se entrelaçam na formação de

bases de dados enviesadas em sistemas de reconhecimento facial. O segundo analisa o viés de gênero no recrutamento automatizado e nos modelos de linguagem, mostrando como algoritmos transformam desigualdades históricas em decisões computacionais. O terceiro desloca o foco da representação distorcida para a invisibilidade seletiva, examinando a exclusão de identidades LGBTQIA.

Juntos, esses casos demonstram o ciclo de retroalimentação da desigualdade algorítmica, em que os dados que estruturam os sistemas reforçam as hierarquias que pretendem representar.

Finalmente, a contextualização nacional considera a realidade específica do Brasil e da África, demonstrando como esses vieses são refratados e amplificados por desigualdades locais. Juntos, esses casos formam a base empírica para construir a tese central sobre o ciclo de retroalimentação da desigualdade algorítmica, no qual os próprios dados que estruturam os sistemas reforçam as hierarquias que pretendem representar.

2.2 O viés interseccional em sistemas de reconhecimento facial

O primeiro caso de estudo, inspirado na pesquisa seminal de Joy Buolamwini e Timnit Gebru (2018), analisa o desempenho de três sistemas comerciais de reconhecimento facial — IBM Watson, Microsoft Azure e Face++ — amplamente utilizados em segurança, autenticação e monitoramento urbano. O estudo revelou taxas de erro superiores a 34% para mulheres negras, enquanto para homens brancos o erro não ultrapassava 1%.

Esses resultados expuseram a natureza estruturalmente racializada das bases de dados utilizadas para treinar os algoritmos. Como o conjunto de imagens era composto majoritariamente por rostos brancos e masculinos, o modelo aprendeu a associar “neutralidade” e “precisão” a essas características fenotípicas, transformando o sujeito branco em uma espécie de “padrão ontológico” da inteligência artificial.

Essa discrepância técnica reflete uma dimensão epistemológica e política: o sistema não apenas reproduz desigualdades, mas as legitima ao revesti-las de objetividade matemática. A falha, portanto, não é apenas de amostragem, mas de racionalidade algorítmica, que naturaliza as hierarquias raciais e de gênero na própria codificação do mundo.

Buolamwini (2018) argumenta que essa assimetria se conecta a dinâmicas mais amplas de vigilância racializada, em que corpos negros e femininos são mais propensos à detecção, monitoramento e suspeição. O problema, portanto,

transcende a dimensão técnica: ele opera como uma forma contemporânea de “racismo algorítmico”, expressão retomada por Ruha Benjamin (2019) ao conceituar o “novo código Jim Crow” — um regime digital que reproduz as lógicas de segregação e exclusão sob a aparência de neutralidade tecnológica.

2.3 A codificação do viés de gênero: do recrutamento aos modelos de linguagem

O segundo caso de estudo examina o viés de gênero em duas camadas interconectadas: (1) a automação de processos de recrutamento corporativo, e (2) a infraestrutura linguística dos grandes modelos de processamento de linguagem natural (PLN).

A primeira camada refere-se ao caso do sistema de recrutamento automatizado da Amazon, revelado em 2018 (Dastin, 2018). O algoritmo, desenvolvido para otimizar a seleção de currículos, foi treinado com dados históricos da própria empresa — um corpus formado majoritariamente por candidatos homens das áreas de engenharia e tecnologia. Como consequência, o sistema aprendeu a penalizar currículos que continham palavras como “women’s” ou “feminist”, e a favorecer descritores associados ao universo masculino, reforçando padrões de exclusão de gênero já consolidados na estrutura corporativa.

Esse caso é emblemático porque evidencia como os dados refletem a cultura organizacional, e como, ao serem transformados em modelo, cristalizam o passado discriminatório como regra de decisão futura. Assim, o algoritmo não apenas aprendeu com a desigualdade — ele a reproduziu e a amplificou, tornando-a computacionalmente eficiente.

A segunda camada aprofunda o problema na base infraestrutural dos modelos de linguagem, como Word2Vec, GloVe e GPT. Esses modelos, treinados em grandes volumes de texto da internet, internalizam padrões linguísticos sexistas: estudos mostram que o modelo Word2Vec associa “homem” a “médico” e “mulher” a “enfermeira”; “homem” a “programador” e “mulher” a “secretária” (Bolukbasi *et al.*, 2016). Essas associações estatísticas, embora matematicamente neutras, traduzem uma semântica social enviesada, perpetuando hierarquias simbólicas em tarefas como tradução automática, geração de texto e análise de sentimentos.

Como observa Safiya Noble (2018), esses sistemas transformam desigualdades históricas em produtos de mercado, consolidando o que ela chama de “algoritmos de opressão”. A questão central, portanto, não é apenas o que a máquina “aprende”, mas o que ela decide esquecer: a pluralidade de experiências e

identidades que não se enquadram no padrão dominante.

Assim, tanto no caso do recrutamento automatizado quanto nos modelos de linguagem, observa-se a emergência de um viés estrutural de gênero que atravessa as etapas de coleta, curadoria, modelagem e aplicação de dados. Esse fenômeno exemplifica como o viés não é um erro técnico isolado, mas uma manifestação estatística da desigualdade social, ancorada em epistemologias excluídas.

2.4 O Poder de Não Contar: A Invisibilidade Seletiva como Mecanismo de Exclusão

Diferenciando-se dos casos de representação distorcida, este foca no fenômeno da invisibilidade seletiva. Investiga-se como a falha estrutural na coleta de dados sobre orientação sexual e identidade de gênero (SOGI) em censos, prontuários eletrônicos de saúde e sistemas de serviços públicos torna a população LGBTQIA+ efetivamente "invisível" para a análise de dados e, conseqüentemente, para os sistemas de IA. As consequências dessa lacuna são graves, indo desde a alocação inadequada de recursos de saúde até a incapacidade de modelos algorítmicos detectarem padrões de discriminação, discurso de ódio e violência direcionados a essa comunidade. Este caso permite uma discussão aprofundada sobre a governança de dados, o direito à autoidentificação e a tensão entre a necessidade de visibilidade para garantir direitos e os riscos de vigilância associados à coleta de dados sensíveis.

Essa mesma lógica de apagamento se manifesta em uma dimensão de desenvolvimento e colonialidade, como ilustra o caso da agricultura de precisão na África. Ali, a vasta diversidade de práticas dos pequenos agricultores, a policultura e os saberes agrícolas tradicionais são sistematicamente omitidos das bases de dados, que privilegiam modelos de monocultura do Norte Global. Ao tornar essas realidades "invisíveis", os algoritmos importados falham em fornecer recomendações úteis, podendo agravar a insegurança alimentar e a dependência tecnológica, em uma clara dinâmica de extrativismo de dados e imposição de modelos epistêmicos (Ajala; Nyerere, 2022).

2.5 Delineamento Metodológico da Análise de Casos

A análise dos casos deste estudo pautou-se por uma abordagem de análise de conteúdo qualitativa, aplicada a um corpus de fontes secundárias — artigos científicos, relatórios técnicos, documentos institucionais e materiais de divulga-

ção científica. Cada caso foi tratado como uma unidade empírica de análise, na qual se buscou identificar padrões discursivos, epistemológicos e políticos relacionados às diferentes manifestações do viés algorítmico.

A seleção dos casos seguiu critérios metodológicos deliberados, fundamentados na literatura sobre estudos de caso (Yin, 2018):

1. Relevância Canônica: escolha de casos amplamente reconhecidos na literatura acadêmica como exemplares na exposição do problema da invisibilidade e da discriminação algorítmica — por exemplo, o estudo *Gender Shades* (Buolamwini; Gebru, 2018) no contexto de reconhecimento facial;
2. Poder Ilustrativo: seleção de casos que demonstram, de forma clara e didática, uma faceta específica da desigualdade algorítmica, como o sistema de recrutamento automatizado da Amazon (Dastin, 2018) e os modelos de linguagem que reproduzem estereótipos de gênero (Bolkvasi *et al.*, 2016);
3. Pertinência Contextual: inclusão de casos empíricos situados em contextos geopolíticos distintos, como o uso de reconhecimento facial na segurança pública brasileira (Ramiro; Cruz, 2023) e a aplicação de IA na agricultura africana sob o paradigma extrativista (Birhane, 2020; Iyer; Achieng; Borokini, 2021).

Cada estudo foi interpretado à luz de um referencial teórico específico — Estudos Críticos de Algoritmos, Feminismo de Dados e Teoria Pós/Decolonial — permitindo uma triangulação entre os níveis técnico, discursivo e sociopolítico. A análise concentrou-se na reconstrução dos mecanismos de produção e legitimação do viés, examinando as estratégias de naturalização, exclusão ou invisibilização embutidas nas práticas algorítmicas.

As fontes para a análise de cada caso foram sistematicamente coletadas de bases de dados acadêmicas (ACM Digital Library, IEEE Xplore, SciELO), relatórios técnicos de institutos de pesquisa (e.g., *AI Now Institute*, ITS Rio) e organizações da sociedade civil, anais de conferências de alto impacto (e.g., ACM FAccT). O procedimento de análise consistiu em: (i) identificar a manifestação central do viés ou da invisibilidade em cada caso; (ii) mapear as consequências sociais, éticas e políticas documentadas nas fontes; e (iii) articular criticamente os achados com os conceitos dos referenciais teóricos adotados. Uma limitação inerente a

esta abordagem é a não realização de auditorias algorítmicas primárias, dependendo, portanto, da validade e do rigor das fontes secundárias analisadas

Em síntese, a análise comparada dos casos permitiu mapear três dimensões convergentes das consequências sociais e políticas do viés algorítmico: (a) a invisibilização epistêmica, em que práticas e saberes locais são apagados pelos modelos de IA; (b) a seletividade estrutural, que reforça desigualdades de gênero, raça e classe por meio da automatização de decisões; e (c) a dependência tecnológica, associada ao extrativismo de dados e à importação acrítica de sistemas.

A articulação entre os achados empíricos e os referenciais teóricos confirma que os Estudos Críticos de Algoritmos explicam o papel do poder e da governança técnica; o Feminismo de Dados evidencia as hierarquias de representação; e a abordagem Pós/Decolonial revela os padrões de dominação e dependência tecnológica global. Juntas, essas lentes permitem compreender a desigualdade algorítmica como um fenômeno multinível — técnico, epistêmico e político.

3 Análise e Desenvolvimento da Argumentação Central

A apresentação dos estudos de caso na seção anterior fornece a base empírica para a análise aprofundada que se segue. Superando a descrição individual de cada caso, esta seção se dedica a sintetizar os achados e a conectá-los para desenvolver a argumentação central deste artigo.

Os resultados evidenciam que as falhas observadas nos sistemas de IA não correspondem a erros técnicos isolados, mas a manifestações recorrentes de lógicas sociais, políticas e históricas de exclusão. Essa constatação emerge de forma transversal em todos os casos analisados: no viés interseccional dos sistemas de reconhecimento facial, na reprodução de estereótipos de gênero pelos modelos de linguagem e no extrativismo digital aplicado à agricultura africana.

Tais fenômenos se materializam em duas formas interligadas de injustiça algorítmica:

1. A representação distorcida, que automatiza e legitima preconceitos históricos sob a aparência de neutralidade técnica;
2. A invisibilidade seletiva, que apaga ou invalida sujeitos, práticas e saberes que não se encaixam nas ontologias dominantes dos sistemas de IA.

Essas dimensões convergem para sustentar a tese de que a IA atua como tecnologia de poder e de governo algorítmico, produzindo e reproduzindo hie-

rarquias sociais por meio de processos de classificação e decisão automatizada. Assim, os estudos de caso demonstram empiricamente que o problema da IA contemporânea é menos uma questão de eficiência técnica e mais uma expressão política e epistêmica da desigualdade.

Os achados empíricos dialogam diretamente com os Estudos Críticos de Algoritmos (Pasquale, 2015; Amore, 2020), que revelam como as infraestruturas digitais se consolidam como sistemas de governança opacos, capazes de regular condutas e distribuir oportunidades de forma invisível. O caso do reconhecimento facial, ao expor a seletividade racial e o uso punitivo da IA, exemplifica essa lógica de governo algorítmico.

À luz do Feminismo de Dados (D'ignázio; Klein, 2020; Noble, 2018), os vieses de gênero nos modelos de linguagem mostram que as tecnologias reencenam desigualdades históricas de representação, evidenciando a necessidade de curadorias de dados situadas e práticas de design sensíveis à diversidade epistemológica.

Por fim, a abordagem Pós/Decolonial (Quijano, 2005; Mbembe, 2019; Birhane, 2020) permite compreender o extrativismo de dados como uma continuidade das estruturas coloniais de exploração — agora mediadas por fluxos de informação e algoritmos que transformam conhecimento local em recurso global. Sob essa lente, a IA deixa de ser vista como instrumento neutro de desenvolvimento e passa a ser entendida como artefato sociotécnico que reproduz dependências geopolíticas e epistemológicas.

Em conjunto, essas três perspectivas evidenciam que os algoritmos não apenas calculam o mundo, mas o normatizam, reconfigurando as fronteiras entre o que é visível e o que é descartável. A análise dos casos, portanto, demonstra que a governança algorítmica contemporânea é um projeto de poder distribuído e automatizado, cujas implicações éticas e políticas exigem escrutínio público, transparência institucional e resistência epistêmica.

3.1 A materialização do viés: da representação distorcida à discriminação automatizada

Os casos de reconhecimento facial e de recrutamento algorítmico ilustram de forma contundente o mecanismo da representação distorcida. A pesquisa seminal de Buolamwini e Gebru (2018) não revela apenas que os sistemas erram, mas que erram de forma sistemática e previsível ao longo de eixos interseccionais de raça e gênero. A alta taxa de erro na identificação de mulheres

negras demonstra que as bases de dados utilizadas para o treinamento desses sistemas foram construídas em torno de um "sujeito padrão" – branco e masculino. Consequentemente, o sistema não apenas falha em reconhecer outros rostos, mas codifica a marginalização desses grupos, tratando-os como um desvio da norma.

Essa mesma lógica é transposta do campo da identificação biométrica para o da oportunidade econômica no caso do sistema de recrutamento da Amazon (Dastin, 2018). O algoritmo não aprendeu a identificar "bons profissionais", mas sim a replicar os padrões de contratação passados da empresa, que eram historicamente enviesados em favor de homens. Ao penalizar currículos com termos associados ao feminino, o sistema automatizou uma barreira de gênero, conferindo ao preconceito uma aparência de objetividade e eficiência técnica. A análise dos modelos de linguagem (Bolukbasi *et al.*, 2016) aprofunda essa constatação, revelando que o viés está gravado na própria infraestrutura semântica da IA, garantindo sua perpetuação em uma miríade de aplicações futuras.

A análise conjunta desses casos permite argumentar que a representação distorcida em bases de dados é o vetor pelo qual a discriminação social é traduzida em código e posta em operação em larga escala. O resultado é a criação de sistemas que não apenas espelham, mas ativamente reforçam e legitimam desigualdades estruturais, tornando o acesso a direitos, serviços e oportunidades condicionado à conformidade com padrões de dados hegemônicos.

3.2 O poder de não contar: a invisibilidade seletiva como mecanismo de exclusão

À luz do Feminismo de Dados (D'ignázio; Klein, 2020) e dos Estudos Críticos de Algoritmos (Pasquale, 2015; Amoore, 2020), a invisibilidade seletiva pode ser compreendida como um dispositivo de poder informacional: ela não apenas omite, mas produz ausências, determinando quem é reconhecido e quem permanece fora dos horizontes da decisão algorítmica.

Se a representação distorcida discrimina quem está nos dados, a invisibilidade seletiva exclui quem sequer chega a ser contado. O estudo sobre a lacuna de dados da população LGBTQIA+ (Hatzenbuehler; Lattanner; Mcketta; Pachankis, 2024) é paradigmático nesse sentido. A ausência de coleta de informações sobre orientação sexual e identidade de gênero em sistemas de saúde e censos não é uma omissão neutra, mas um ato político de apagamento, que remove as necessidades dessa comunidade do mapa das políticas públicas e dos sistemas de

IA informados por dados. Para o algoritmo, o que não está presente no conjunto de treinamento simplesmente não existe. Essa invisibilidade tem consequências diretas: má alocação de recursos, falhas na detecção de disparidades de saúde e incapacidade de plataformas digitais em proteger usuários LGBTQIA+ de ataques sistemáticos (Scheuerman *et al.*, 2021).

Sob uma lente pós-colonial (Quijano, 2005; Mbembe, 2019), essa lógica de exclusão se repete na contextualização nacional do reconhecimento facial na segurança pública brasileira (Rede De Observatórios Da Segurança, 2021). Além da representação distorcida, que produz erros desproporcionais contra a população negra, há aqui um componente epistêmico de invisibilidade. A importação acrítica de tecnologias treinadas com dados do Norte Global invisibiliza a complexidade racial e cultural brasileira, promovendo um “transplante tecnológico” que falha em reconhecer corpos miscigenados e contextos locais de vulnerabilidade. Nessa perspectiva, a invisibilidade é uma forma de colonialidade algorítmica, em que a realidade nacional é reinterpretada segundo parâmetros externos que não a contemplam.

A análise desses casos fundamenta o argumento de que a decisão sobre o que e quem contar é, em si, um exercício de poder. Contar é reconhecer, e deixar de contar é uma forma contemporânea de marginalização. A invisibilidade nos dados nega a existência, as necessidades e os direitos de populações inteiras, deixando-as à margem de uma sociedade cada vez mais governada por lógicas algorítmicas e pela economia política da atenção e da visibilidade. Assim, a exclusão algorítmica opera não apenas pela discriminação explícita, mas também — e talvez de forma mais profunda — pela produção sistemática do esquecimento.

3.3 Síntese da argumentação central: o ciclo de retroalimentação da desigualdade algorítmica

A análise integrada dos estudos de caso revela que a representação distorcida e a invisibilidade seletiva não são fenômenos independentes, mas etapas interligadas de um mesmo processo: o ciclo de retroalimentação da desigualdade algorítmica. Esse ciclo descreve como estruturas sociais de exclusão são traduzidas em dados, processadas por modelos de Inteligência Artificial e, finalmente, devolvidas à sociedade sob a aparência de neutralidade técnica.

Na primeira fase, ocorre a absorção das hierarquias sociais. Os dados de treinamento refletem as desigualdades históricas de raça, gênero, classe e território. Essa assimetria inicial define o que será reconhecido como relevante e

quem será incluído ou excluído da codificação algorítmica.

Na segunda fase, o processamento algorítmico transforma essas distorções em padrões estatísticos. Desprovidos de contexto social, os modelos interpretam desigualdades como regularidades legítimas. O preconceito é reconfigurado em lógica matemática — o que antes era discriminação torna-se correlação.

A terceira fase é a aplicação prática desses modelos em domínios sensíveis como segurança, crédito, saúde e trabalho. As decisões enviesadas passam a operar em larga escala, com aparência de eficiência e objetividade, legitimando desigualdades pré-existentes e criando novas barreiras invisíveis à contestação.

Por fim, a quarta fase, de retroalimentação, completa o ciclo. Os resultados produzidos — como prisões indevidas, negação de crédito ou exclusão em processos seletivos — geram novos dados que confirmam e reforçam as premissas originais do sistema. Assim, a desigualdade não apenas se perpetua, mas se autojustifica, criando uma forma de governo algorítmico em que o poder se distribui por infraestruturas automatizadas e opacas.

Esse modelo evidencia que interromper o ciclo requer mais do que correções técnicas de “viés”. É necessário politizar o debate sobre dados, questionando as estruturas de poder que definem quem é representado, como é representado e com que finalidade. Somente ao desestabilizar esses fundamentos — epistemológicos, econômicos e coloniais — é possível evitar que a Inteligência Artificial continue a transformar desigualdades históricas em infraestruturas do esquecimento.

3.4 A colonialidade dos dados: o horizonte global do governo algorítmico

O ciclo descrito na Figura 01 não opera em um vácuo técnico, mas está inserido em infraestruturas globais de produção de dados que reproduzem antigas assimetrias coloniais. À luz da abordagem pós/decolonial (Quijano, 2005; Mbembe, 2019; Birhane, 2020), pode-se compreender a Inteligência Artificial como parte de um regime planetário de governo algorítmico, em que o poder é exercido por meio do controle sobre fluxos informacionais, plataformas e padrões de interoperabilidade.

Essa dinâmica produz o que Pasquale (2015) chama de “sociedade da caixa-preta”: sistemas opacos, baseados em dados coletados de populações periféricas, operados por centros hegemônicos e regulados por métricas de eficiência empresarial. Assim, o Sul Global torna-se fornecedor de dados e mão de obra digital, enquanto o Norte Global concentra o poder de modelar e validar o conhe-

cimento (Noble, 2018; D'ignázio; Klein, 2020).

A colonialidade dos dados é, portanto, o horizonte mais amplo do ciclo da desigualdade algorítmica: ela conecta o nível técnico às lógicas históricas de exploração e dependência. Sob o discurso da inovação, a IA converte saberes locais em insumos estatísticos, despolitizando o processo de dominação. Reconhecer essa estrutura é essencial para que políticas públicas e práticas científicas possam reverter a assimetria epistêmica e construir uma Inteligência Artificial verdadeiramente situada, plural e democrática.

3.5 Os Limites do "Debiasing": Uma Crítica às Soluções Técnicas

A constatação do ciclo de retroalimentação da desigualdade demanda uma análise crítica das soluções técnicas propostas no campo da IA, frequentemente agrupadas sob o rótulo de "fairness" (justiça) e "debiasing" (correção de viés) (Mehrabi *et al.*, 2021). Embora meritórias, essas abordagens apresentam limitações conceituais e práticas que reforçam a necessidade de uma crítica mais estrutural (Ferretti; Montibeller; Von Winterfeldt, 2023).

O campo computacional desenvolveu múltiplas métricas formais de justiça. Por exemplo, a Paridade Demográfica postula que a proporção de indivíduos que recebem um determinado resultado positivo (e.g., a aprovação de um empréstimo) deve ser a mesma entre diferentes grupos demográficos (Denis *et al.*, 2024). Contudo, esta abordagem pode levar à violação de princípios meritocráticos, ao forçar a seleção de indivíduos menos qualificados para atingir uma cota estatística. Em contrapartida, a Igualdade de Oportunidade exige que as taxas de verdadeiros positivos e falsos positivos sejam iguais entre os grupos, focando na equidade do tratamento para indivíduos igualmente qualificados (Hardt; Price; Srebro, 2016).

O problema central, documentado na literatura técnica, é a impossibilidade matemática de satisfazer simultaneamente múltiplas dessas métricas de justiça na maioria dos cenários do mundo real. A escolha de otimizar para a Paridade Demográfica pode piorar a Igualdade de Oportunidade, e vice-versa. Isso revela que a "justiça" algorítmica não é um problema de otimização técnica, mas um campo de decisões normativas e políticas sobre qual definição de justiça priorizar em um determinado contexto. É nesse ponto que a crítica ao "ethics-washing" se aprofunda: uma organização pode adotar a métrica que lhe é mais conveniente ou que menos afeta seu desempenho, declarando seu sistema "justo" com base nesse critério específico, enquanto ignora as injustiças produzidas

sob outras óticas. Essa abordagem mascara a natureza política da escolha sob um verniz de conformidade técnica, sem alterar as estruturas de poder que produzem a desigualdade em primeiro lugar.

3.6 Diretrizes para a justiça algorítmica: uma perspectiva sociotécnica e decolonial

A superação do ciclo de retroalimentação da desigualdade algorítmica exige uma transição paradigmática: de uma abordagem reativa de “correção de vieses” para uma busca proativa e contínua por justiça algorítmica (Rosa; Guasque, 2024). Por conseguinte, essa justiça não pode ser alcançada por meio de otimizações técnicas isoladas, mas requer intervenções estruturais, procedimentais e epistemológicas em toda a ecologia da IA. Nesse sentido, propomos um conjunto de diretrizes interdependentes, articuladas nos eixos dos dados, dos algoritmos e da governança, que visam não apenas mitigar danos, mas, fundamentalmente, reorientar o poder.

Em primeiro lugar, o eixo dos dados demanda uma nova epistemologia informacional. É necessário abandonar a premissa extrativista que os concebe como recurso bruto a ser explorado e adotar práticas de curadoria eticamente orientada. Assim, torna-se essencial a documentação compulsória e padronizada sobre a proveniência, as limitações e a arquitetura dos dados e modelos, através de ferramentas como Datasheets for Datasets (Gebru *et al.*, 2021) e Model Cards for Model Reporting (Mitchell *et al.*, 2019). De modo complementar, é preciso avançar em direção a modelos de soberania de dados comunitária, por meio da criação de cooperativas ou trusts de dados, dotados de personalidade jurídica e capazes de negociar coletivamente os termos de uso das informações de uma comunidade — inclusive vetando aplicações consideradas danosas.

Em segundo lugar, o eixo algorítmico deve traduzir a transparência em contestabilidade efetiva. Portanto, auditorias externas, independentes e contínuas (Raji; Buolamwini, 2019) não devem visar apenas à produção de relatórios, mas à constituição de um verdadeiro “devido processo algorítmico”, com vias de recurso claras e acessíveis para que cidadãos possam contestar decisões automatizadas que os afetem adversamente. Entretanto, a interpretabilidade (XAI) precisa ser abordada criticamente, pois uma explicação superficial pode ofuscar uma lógica estruturalmente injusta. Ademais, a escolha entre diferentes métricas de justiça (Verma; Rubin, 2018) deve ser reconhecida como uma decisão política, que expressa uma teoria particular de justiça e, por isso, deve ser objeto de deli-

beração pública em domínios de alto impacto social.

Por fim, o eixo da governança democrática e regulatória é o que garante a sustentabilidade institucional dessas práticas. Nesse contexto, torna-se crucial estabelecer marcos regulatórios vinculantes que exijam avaliações prévias de impacto em direitos humanos para sistemas de IA de alto risco, definindo regimes claros de responsabilização e reparação. Contudo, para que a regulação seja efetiva, é indispensável fomentar ecossistemas de co-governança — envolvendo Estado, setor privado, academia e sociedade civil — e investir em letramento crítico em IA como instrumento de empoderamento cívico. Em contraste com a adoção acrítica de modelos internacionais, a abordagem decolonial enfatiza a autodeterminação regulatória, de modo que cada contexto desenvolva marcos normativos coerentes com suas realidades sociais e tradições jurídicas.

Considerações finais

Este artigo demonstrou que a designação “viés algorítmico” funciona como um eufemismo que obscurece a natureza intrinsecamente política da automação. Longe de serem falhas técnicas ou anomalias, as exclusões e os danos perpetrados por sistemas de Inteligência Artificial são a consequência lógica de sua inscrição em epistemologias hegemônicas. A análise revelou o ciclo de retroalimentação pelo qual vetores históricos de opressão ganham nova velocidade e escala, sendo codificados na própria arquitetura dos modelos. Trata-se de um mecanismo que transmuta poder social em poder computacional, naturalizando a desigualdade sob um verniz de neutralidade e objetividade técnica.

Em resposta a esse diagnóstico, o trabalho avançou para além da crítica ao apresentar um arcabouço conceitual com diretrizes para a justiça algorítmica. Esse framework, fundamentado em três pilares interdependentes, propõe uma alternativa às soluções puramente tecnocráticas. O primeiro pilar, a soberania dos dados, inspira-se nas teorias de data justice (Heeks; Renken, 2018) e data sovereignty (Carroll *et al.*, 2020), mas amplia-as ao situá-las em um horizonte comunitário e decolonial, no qual as próprias comunidades definem os termos de uso e circulação de seus dados. O segundo pilar, a contestabilidade efetiva, dialoga com a tradição da algorithmic accountability (Citron, 2007; Selbst, 2017), mas desloca o foco da transparência para a capacidade real de contestação, reivindicando um “devido processo algorítmico” análogo às garantias jurídicas clássicas. O terceiro pilar, a governança democrática e regulatória, apoia-se nas abordagens de AI governance e participatory design (Fjeld *et al.*, 2020; Costanza-Chock, 2020),

ao mesmo tempo em que incorpora o alerta da literatura decolonial (Mbembe, 2019; Birhane, 2020) contra a importação acrítica de modelos normativos do Norte Global.

Distintamente de propostas existentes, que permanecem restritas a métricas de fairness ou a instrumentos de conformidade ética, o framework aqui desenvolvido sustenta que a justiça algorítmica é inseparável da redistribuição de poder. As métricas de equidade, embora úteis em contextos limitados, são insuficientes para desmantelar iniquidades e podem, paradoxalmente, operar como estratégias de ethics-washing que legitimam sistemas estruturalmente injustos.

As implicações deste estudo exigem, portanto, uma reorientação da agenda de pesquisa: de uma postura de auditoria reativa para a de imaginação e engenharia de infraestruturas computacionais radicalmente distintas, baseadas em lógicas não extrativistas e participativas. Uma postura crítica consequente deve também fomentar o debate público sobre a necessidade de moratórias ou de não implementação de tecnologias — como o policiamento preditivo e o score social — que se mostram incompatíveis com os alicerces democráticos e os direitos fundamentais. A questão fulcral, portanto, desloca-se: não se trata de otimizar a “justiça” em sistemas de IA, mas de deliberar politicamente sobre quem tem o poder de decidir, e sob quais condições éticas e sociais legitimamos sua existência e operação.

Referências

- ARRATIBEL, A. J. **Paola Ricaurte talks about the future of inclusive AI**. TecScience, june 2025.
- AJALA, F.; NYERERE, I. Digital Harvests, Digital Shadows: the risks of data colonialism in african smallholder agriculture. Nairobi: African Centre for Technology Studies, 2022.
- AMOORE, L. **Cloud Ethics**: algorithms and the attributes of ourselves and others. Durham: Duke University Press, 2020.
- BAROCAS, S.; HARDT, M.; NARAYANAN, A. **Fairness and Machine Learning**: limitations and opportunities. Cambridge: MIT Press, 2019. Disponível em: <https://fairmlbook.org>. Acesso em: 22 out. 2025.
- BENJAMIN, R. **Race After Technology**: abolitionist tools for the new Jim Code. Cambridge: Polity Press, 2019.
- BIRHANE, A. Algorithmic Colonization of Africa. **SCRIPTed**, v. 17, n. 2, p. 389-405, 2020. Disponível em: <https://script-ed.org/article/algorithmic-colonization-of-africa/>. Acesso em: 23 out. 2025.

BOLUKBASI, T.; CHANG, K-W.; ZOU, J.; SALIGRAMA V.; KALAI A. Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings. In: **Conference On Neural Information Processing Systems**, 30., 2016, Barcelona. p. 4349–4357, 2016. Disponível em: <https://papers.nips.cc/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html>. Acesso em: 22 out. 2025.

BUOLAMWINI, J.; GEBRU, T. Gender Shades: intersectional accuracy disparities in commercial gender classification. In: **Conference On Fairness, Accountability, And Transparency**, 2018, New York. Proceedings [...]. Cambridge: PMLR, 2018. v. 81, p. 77-91.

CARROLL, S. R.; GARBA, I.; FIGUEROA-RODRÍGUEZ, O. L.; HOLBROOK, J.; LOVETT, R.; MATE-RECHERA, S.; PARSONS, M.; RASEROKA, K.; RODRIGUEZ-LONEBEAR, D.; ROWE, R.; SARA, R.; WALKER, J. D.; ANDERSON, J.; HUDSON, M. The CARE principles for indigenous data governance. **Data science journal**, v. 19, p. 2-12 2020. Disponível em: <https://datascience.codata.org/articles/1158/files/submission/proof/1158-1-8528-2-10-20201104.pdf>. Acesso em: 28 out. 2025.

CARVALHO, A. A. de; BARRETO, R. C. V. A invisibilidade das pessoas LGBTQIA+ nas bases de dados: novas possibilidades na Pesquisa Nacional de Saúde 2019?. **Ciência & Saúde Coletiva**, v. 26, p. 4059-4064, 2021. Disponível em: <https://www.scielo.br/j/csc/a/rwDkNhDCdyY5xdfyXNxmm-GH/?format=html&lang=pt>. Acesso em: 28 out. 2025.

CITRON, D. K. Technological due process. **Wash. UL Rev.**, v. 85, p. 1249, 2007. Disponível em: https://openscholarship.wustl.edu/cgi/viewcontent.cgi?article=1166&context=law_lawreview. Acesso em: 28 out. 2025.

COSTANZA-CHOCK, S. **Design justice**: community-led practices to build the worlds we need. The MIT Press, 2020.

COULDRY, N.; MEJIAS, U. A. **The Costs of Connection**: how data is colonizing human life and appropriating it for capitalism. Stanford: Stanford University Press, 2019.

CRAWFORD, K. **Atlas of AI**: power, politics, and the planetary costs of artificial intelligence. New Haven: Yale University Press, 2021.

DASTIN, J. Amazon scraps secret AI recruiting tool that showed bias against women. **Reuters**, 10 out. 2018. Disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>. Acesso em: 7 jul. 2025.

DE LIMA VIANA, G. M.; SPERANDEO DE MACEDO, C. Inteligência artificial e a discriminação algorítmica: : uma análise do caso amazon. **Direito & TI**, [S. l.], v. 1, n. 19, p. 39–62, 2024. DOI: 10.63451/ti.v1i19.212. Disponível em: <https://www.direitoeti.com.br/direitoeti/article/view/212>. Acesso em: 4 nov. 2025.

DENIS, C.; ELIE R.; HEBIRI, M.; HU F. Fairness guarantees in multi-class classification with demographic parity. **Journal of Machine Learning Research**, v. 25, n. 130, p. 1- 46, 2024. Disponível em: <https://arxiv.org/abs/2109.13642>. Acesso em: 28 out. 2025.

D'IGNAZIO, C.; KLEIN, L. F. **Data Feminism**. Cambridge: The MIT Press, 2020.

FERRETTI, V.; MONTIBELLER, G.; VON WINTERFELDT, D. Testing the effectiveness of debiasing techniques to reduce overprecision in the elicitation of subjective continuous probability distributions. **European Journal of Operational Research**, v. 304, n. 2, p. 661-675, 2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0377221722003046>. Acesso em 28 out. 2025.

FJELD, J.; ACHTEN, N.; HILLIGOSS, H.; NAGY, A. C.; SRIKUMAR, M. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. **Berkman Klein**

Center Research Publication, n. 2020-1, 2020. Disponível em: https://www.researchgate.net/publication/339138141_Principled_Artificial_Intelligence_Mapping_Consensus_in_Ethical_and_Rights-Based_Approaches_to_Principles_for_AI. Acesso em: 28 out. 2025.

GEBRU, T.; MORGENSTERN, J.; VECCHIONE B.; VAUGHAN, J. W. WALLACH, H.; DAUMÉ III, H.; CRAWFORD, K. Datasheets for Datasets. **Communications of the ACM**, New York, v. 64, n. 12, p. 86–92, dez. 2021. Disponível em: <https://arxiv.org/abs/1803.09010>. Acesso em: 28 out. 2025.

GUPTA, N.; MUJUMDAR, S.; PATEL, H.; MASUDA, S.; PANWAR, N.; BANDYOPADHYAY, S.; MEHTA, S.; GUTTULA, S.; AFZAL, S.; MITTAL, R. S.; MUNIGALA, V. Data quality for machine learning tasks. In: **Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining**. The ACM Digital Library. KDD '21, p. 4040–4041. 2021. Disponível em: <https://dl.acm.org/doi/abs/10.1145/3447548.3470817>. Acesso em: 30 out. 2025.

HARDT, M.; PRICE, E.; SREBRO, N. Equality of opportunity in supervised learning. In: **Proceedings of the 30th International Conference on Neural Information Processing Systems**. v. 29, p. 3323–3331, 2016. Disponível em: <https://dl.acm.org/doi/abs/10.5555/3157382.3157469>. Acesso em: 30 out. 2025.

HATZENBUEHLER, M. L.; LATTANNER, M. R.; MCKETTA, S.; PACHANKIS, J. E. Structural stigma and LGBTQ+ health: a narrative review of quantitative studies. *The Lancet Public Health*, [S.L.], v. 9, n. 2, p. 109–127, fev. 2024. Elsevier BV. [http://dx.doi.org/10.1016/s2468-2667\(23\)00312-2](http://dx.doi.org/10.1016/s2468-2667(23)00312-2). Disponível em: <https://www.thelancet.com/action/showPdf?pii=S2468-2667%2823%2900312-2>. Acesso em: 04 nov. 2025. YIN, R. K. **Case Study Research and Applications: design and methods**. 6. ed. Thousand Oaks: Sage, 2018.

HEEKS, R.; RENKEN, J. Data justice for development: What would it mean? **Information Development**, v. 34, n. 1, p. 90–102, 2018. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/0266666916678282>. Acesso em: 30 out. 2025.

IYER, N.; ACHIENG, G.; BOROKINI, F.; LUDGER, U. **Automated imperialism, expansionist dreams: exploring digital extractivism in Africa**. POLICY, 2021. Disponível em: <https://archive.pollicy.org/wp-content/uploads/2021/06/Automated-Imperialism-Expansionist-Dreams-Exploring-Digital-Extractivism-in-Africa.pdf>. Acesso em: 30 out. 2025.

MBEMBE, A. **Necropolítica**. São Paulo: n-1 Edições, 2019.

MEHRABI, N.; MORSTATTER, F.; SAXENA, N.; LERMAN, K.; GALSTYAN, A. A survey on bias and fairness in machine learning. **ACM computing surveys (CSUR)**, v. 54, n. 6, p. 1–35, 2021. Disponível em: <https://dl.acm.org/doi/10.1145/3457607>. Acesso em: 30 out. 2025.

MITCHELL, M.; WU, S.; ZALDIVAR, A.; BARNES, P.; VASSERMAN, L.; HUTCHINSON, B.; SPITZER, E.; RAJI, I. D.; GEGRU, T. Model Cards for Model Reporting. In: **Conference On Fairness, Accountability, And Transparency**, 2019, Atlanta. Proceedings [...]. New York: ACM, 2019. p. 220–229. Disponível em: <https://arxiv.org/abs/1810.03993>. Acesso em: 30 out. 2025.

NOBLE, S. U. **Algorithms of Oppression**: how search engines reinforce racism. New York: New York University Press, 2018.

ORR, W.; CRAWFORD, K. The social construction of datasets: on the practices, processes, and challenges of dataset creation for machine learning. **New Media & Society**, v. 26, n. 9, p. 4955–4972, 2024.

PASQUALE, F. **The Black Box Society**: the secret algorithms that control money and information. Cambridge: Harvard University Press, 2015.

QUIJANO, A. A colonialidade do saber: eurocentrismo e ciências sociais. Perspectivas latino-americanas. Buenos Aires: **CLACSO**, p. 117-142, 2005. Disponível em: <https://ufrb.edu.br/educacao-docampocfp/images/Edgardo-Lander-org-A-Colonialidade-do-Saber-eurocentrismo-e-ciC3A-Ancias-sociais-perspectivas-latinoamericanas-LIVRO.pdf>. Acesso em: 30 out. 2025.

RAJI, I. D.; BUOLAMWINI, J. Actionable Auditing: investigating the impact of publicly naming biased performance results of commercial AI products. In: **AAAI/ACM Conference On Ai, Ethics, And Society**, 2019, Honolulu. Proceedings [...]. New York: ACM, 2019. p. 429-435. Disponível em: <https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/>. Acesso em: 30 out. 2025.

RAMIRO, A.; CRUZ, L. The grey-zones of public-private surveillance: Policy tendencies of facial recognition for public security in Brazilian cities. **Internet Policy Review**, v. 12, n. 1. p. 1-28, 2023. Disponível em: <https://policyreview.info/pdf/policyreview-2023-1-1705.pdf>. Acesso em: 30 out. 2025.

REDE de Observatórios Da Segurança. **Pele Alvo**: a cor da violência policial. [S.l.]: Rede de Observatórios da Segurança, 2021. Disponível em: <https://redeobservatorios.com.br/wp-content/uploads/2022/08/Pele-Alvo-A-cor-da-violencia-policial.pdf>. Acesso em: 7 jul. 2025.

ROCHE, C.; WALL, P. J.; LEWIS, D. Ethics and diversity in artificial intelligence policies, strategies and initiatives. **AI and Ethics**, v. 3, n. 4, p. 1095-1115, 2023. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/36246014/>. Acesso em: 30 out. 2025.

ROSA, A. M. da; GUASQUE, B. Inteligência artificial, vieses algorítmicos e racismo: o lado desconhecido da justiça algorítmica. **Opinión Jurídica**, v. 23, n. 50, p. 1-23, 2024. Disponível em: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-25302024000200008. Acesso em: 30 out. 2025.

SELBST, A. D. Disparate impact in big data policing. **Ga. L. Rev.**, v. 52, p. 109, 2017. Disponível em: https://dlg.usg.edu/record/uglaw_glr_vol52-iss1-6. Acesso em: 30 out. 2025.

SCHEUERMAN, M. K.; JIANG, J. A.; FIESLER, C.; BRUBAKER, J. R. 'What is it about 'lesbian' that is so scary?': How AI-driven content moderation systems fail LGBTQ+ creators. **Proceedings of the ACM on Human-Computer Interaction**, New York, v. 5, n. CSCW2, art. 422, p. 368:1-35, out. 2021. Disponível em: <https://arxiv.org/pdf/2108.04401>. Acesso em: 30 out. 2025.

VERMA, S.; RUBIN, J. Fairness definitions explained. In: IEEE/ACM INTERNATIONAL WORKSHOP ON SOFTWARE FAIRNESS, 2018, Gotemburgo. **Proceedings** [...]. New York: IEEE, 2018. p. 1-7. Disponível em: <https://dl.acm.org/doi/10.1145/3194770.3194776>. Acesso em: 01 nov. 2025.

WANG, A.; RAMASWAMY, V. V.; RUSSAKOVSKY, O. Towards intersectionality in machine learning: including more identities, handling underrepresentation, and performing evaluation. In: **Proceedings of the 2022 ACM conference on fairness, accountability, and transparency**. 2022. p. 336-349. Disponível em: <https://dl.acm.org/doi/fullHtml/10.1145/3531146.3533101>. Acesso em: 01 nov. 2025.

YIN, R. K. **Case Study Research and Applications**: design and methods. 6. ed. Thousand Oaks: Sage, 2018.

Sobre os autores e a autora

Jefferson Igor Duarte Silva - Mestre em Sistemas e Computação. Professor do Instituto Federal do RN, Campus Ipanguaçu. E-mail: jefferson.igorbr@gmail.com. Lattes: <https://lattes.cnpq.br/8083389022423202>. ORCID: <https://orcid.org/0000-0002-4346-1160>.

Akynara Aglaé Rodrigues Santos da Silva Burlamaqui - Doutora em Educação. Professora Adjunta da Universidade Federal Rural do Semiárido (UFERSA). E-mail: akynara@gmail.com. Lattes: <https://lattes.cnpq.br/9702206935610428>. ORCID: <https://orcid.org/0000-0002-8941-9128>.

Aquiles Medeiros Filgueira Burlamaqui - Doutor em Engenharia Elétrica. Professor Associado da Universidade Federal do RN (UFRN). E-mail: aquilesburlamaqui@gmail.com. Lattes: <https://lattes.cnpq.br/8670475877813913>. ORCID: <https://orcid.org/0000-0001-6754-8335>.