PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
Programa de Pós-Graduação em Letras

Jair Alcindo Lobo de Melo

RACISMO ALGORÍTMICO À LUZ DA ANALÍTICA DO BIOPODER DE MICHEL FOUCAULT

Belo Horizonte
2025

Jair Alcindo Lobo de Melo

Racismo algorítmico à luz da analítica do biopoder de Michel Foucault

Texto apresentado para Exame de defesa de Tese ao Programa de Pós-Graduação em Letras da Pontifícia Universidade Católica de Minas Gerais como requisito parcial para a obtenção do título de Doutor em Letras.

Orientadora: Prof.ª Dr.ª Jane Quintiliano Guimarães Silva.

Área de concentração: Linguística e Língua Portuguesa / Linha de pesquisa: Linguagem e Enunciação: Interações Sociais e Práticas Discursivas.

Belo Horizonte
2025

FICHA CATALOGRÁFICA
Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

Melo, Jair Alcindo Lobo de
M528r Racismo algorítmico à luz da analítica do biopoder de Michel Foucault / Jair Alcindo Lobo de Melo. Belo Horizonte, 2023.
194 f. : il.

Orientadora: Jane Quintiliano Guimarães Silva
Tese (Doutorado) - Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Letras

1. Foucault, Michel, 1926-1984 - Crítica e interpretação. 2. Poder (Ciências sociais). 3 Biopolítica. 4. Racismo. 5. Discriminação. 6. Tecnologia da informação e da comunicação. 7. Inteligência artificial. 8. Análise do discurso. I. Silva, Jane Quintiliano Guimarães. II. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Letras. III. Título.

CDU: 800.852

Ficha catalográfica elaborada por Fabiana Marques de Souza e Silva - CRB 6/2086

Jair Alcindo Lobo de Melo

Racismo algorítmico à luz da analítica do biopoder de Michel Foucault

Texto apresentado para Exame de defesa de Tese ao Programa de Pós-Graduação em Letras da Pontifícia Universidade Católica de Minas Gerais como requisito parcial para a obtenção do título de Doutor em Letras.

_____
Prof.ª Dr.ª Jane Quintiliano Guimarães Silva – PUC Minas (Orientadora)

_____
Prof.ª Dr.ª Daniella Lopes Dias Ignácio Rodrigues (PUC Minas)

_____
Prof. Dr. Paulo Henrique Aguiar Mendes (UFOP – Membro Externo)

_____
Prof.ª Dr.ª Eliara Santana Ferreira (Escola do Legislativo ALMG– Membro Externo)

_____
Prof.ª Dr.ª Ayvania Alves Pinto (CISEB/SEDUCPA – Membro Externo)

Belo Horizonte, 12 de agosto de 2025

AGRADECIMENTOS

À minha orientadora, escuta qualificada e parceira neste trabalho, Jane Quintiliano Guimarães Silva, pelas orientações virtuais, mas não menos acaloradas e calorosas pelo tema e pelos autores envolvidos, pela confiança e pelo carinho com que me conduziu nesse terreno movediço chamado Michel Foucault.

Aos docentes do Programa de Pós-Graduação em Letras da PUC-MG, pelos ensinamentos, mesmo diante de uma pandemia que os condicionou a ressignificar a forma de ensinar e aprender.

À Prof.ª Dr.ª Daniella Lopes Dias Ignácio Rodrigues Lopes (PUC Minas), à Prof.ª Dr.ª Eliara Santana Ferreira (UFMG), à Prof.ª Dr.ª Ayvania Alves Pinto e ao Prof. Dr. Paulo Henrique Aguiar Mendes, integrantes banca avaliadora desta tese, cujas contribuições me auxiliaram, sobremaneira, a chegar a este formato do texto.

À minha esposa Roberta Rios e aos meus filhos Matheus, Davi, Lucca e Alice, que, apesar de terem presenciado tempos de angústia e ausências, sempre foram parceiros ao longo da minha jornada nesse metaverso chamado Doutorado.

O poder não é algo que se possui, mas algo que se exerce.
Michel Foucault.

RESUMO

Esta pesquisa investiga as formas pelas quais o racismo algorítmico se manifesta nas tecnologias digitais, explorando as estruturas de poder e os discursos que sustentam e perpetuam práticas discriminatórias. O problema central consiste em compreender como os sistemas automatizados, embora frequentemente apresentados como neutros, reproduzem e amplificam preconceitos raciais presentes na sociedade, impactando decisivamente a vida de minorias raciais. Esta tese investiga as práticas discursivas racializadas e as suas relações com os mecanismos de produção de vieses nos algoritmos, além de compreender como o biopoder, na perspectiva de Foucault (1998, 2008), se manifesta nesse contexto digital. Nesse campo, além do conceito de biopoder, do teórico francês também são tomados os conceitos de discurso e, na esteira de Pêcheux (2010), formações discursivas e seus efeitos na configuração de subjetividades e estruturas sociais. A pesquisa também adota como suporte a contribuição de estudiosos de outros campos do saber, como Benjamin (2019), Carneiro (2005), Collins (2019), Da Silva (2020), Eubanks (2018), Han (2022), Mbembe (2022), Munanga (1999), Zuboff (2021), entre outros, que permitem a compreensão crítica do impacto das tecnologias digitais na perpetuação de desigualdades e incentivam uma reflexão ética e política sobre o papel do poder algorítmico na sociedade atual. Metodologicamente, a pesquisa realiza uma abordagem arqueológica das práticas discursivas enviesadas presentes nas tecnologias, examinando os discursos e as ações institucionais que sustentam os algoritmos e suas decisões automatizadas. Quanto aos resultados, evidenciam que os sistemas algorítmicos, muitas vezes considerados neutros, carregam e reproduzem valores racistas, institucionalizando uma opressão racializada que reforça desigualdades sociais, deixando patente que o racismo algorítmico se constitui como uma nova formação discursiva.

Palavras-chave: Racismo Algorítmico, Formação Discursiva, Poder, Biopoder, Biopolítica.

ABSTRACT

This research investigates the ways in which algorithmic racism manifests itself in digital technologies, exploring the power structures and discourses that sustain and perpetuate discriminatory practices. The central problem is understanding how automated systems, although often presented as neutral, reproduce and amplify racial prejudices present in society, decisively impacting the lives of racial minorities. This thesis investigates racialized discursive practices and their relationship with the mechanisms of bias production in algorithms, in addition to understanding how biopower, from Foucault's perspective (1998, 2008), manifests itself in this digital context. In this field, in addition to the French theorist's concept of biopower, the concepts of discourse are also taken from the concept of discourse and, following Pêcheux (2010), discursive formations and their effects on the configuration of subjectivities and social structures. The research also draws on contributions from scholars from other fields, such as Benjamin (2019), Carneiro (2005), Collins (2019), Da Silva (2020), Eubanks (2018), Han (2022), Mbembe (2022), Munanga (1999), and Zuboff (2021), among others. These contributions allow for a critical understanding of the impact of digital technologies on perpetuating inequalities and encourage ethical and political reflection on the role of algorithmic power in today's society. Methodologically, the research takes an archaeological approach to the biased discursive practices present in these technologies, examining the discourses and institutional actions that support algorithms and their automated decisions. The results demonstrate that algorithmic systems, often considered neutral, carry and reproduce racist values, institutionalizing racialized oppression that reinforces social inequalities, clearly demonstrating that algorithmic racism constitutes a new discursive formation.

Keywords: Algorithmic Racism, Discursive Formation, Power, Biopower, Biopolitics.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO

Nas últimas décadas, o avanço tecnológico e o uso de algoritmos[1] em processos decisórios, plataformas digitais, mecanismos de busca, mídias sociais e inúmeras outras ações cotidianas transformaram significativamente nossa sociedade. Cabe salientar que a promessa de neutralidade e eficiência associada a essas novas tecnologias nos ambientes digitais atuais esconde, contudo, a realidade de que esses sistemas, ao serem alimentados por dados históricos e sociais já instituídos e consolidados, fortalecem as relações e os modos de dizer, os quais reiteram um discurso que não é novo, mas que se atualiza na contemporaneidade, que reproduz e, em alguns casos, expande desigualdades e discriminações preexistentes.

Assim sendo, pode-se reiterar que a sociedade contemporânea tem vivenciado expressivos avanços no desenvolvimento de algoritmos e sistemas de inteligência artificial, aplicados largamente em várias áreas, desde a seleção de currículos até o sistema de justiça criminal. No entanto, há crescentes preocupações sobre o modo como esses vieses estão fortalecendo e sedimentando estruturas racistas e discriminatórias, o que resulta em desigualdades sistemáticas para grupos raciais e outros minoritários. Por meio de alguns conceitos que explicitam as problemáticas que esses sistemas têm representado para a ampliação das desigualdades raciais, teóricos e ativistas da área estão contribuindo para o alargamento do campo de estudos, com discussões e reflexões acerca das roupagens que o racismo, em particular, por ser foco desta pesquisa, possui na contemporaneidade.

Dentre essas mudanças significativas na experiência racial em meios digitais, emerge a questão de que a discriminação racial está embutida na programação de sistemas, e isso aumenta seus riscos e suas consequências à medida que o uso cotidiano dessas tecnologias se assenta nos mais diversos setores. Nas tecnologias de inteligência artificial, que, em sua maioria, são formuladas para desenvolver parte de suas ações de forma automática, a discriminação pode ser ainda mais grave. Ela está lá, querendo ou não, o que tornou o campo de estudos e de discussões jurídicas

[1] Algoritmos são conjuntos de regras e processos lógicos utilizados para resolver problemas ou tomar decisões. Eles estão presentes em sistemas de recomendação, reconhecimento facial, análises de crédito, policiamento preditivo e até em processos seletivos. Apesar de sua aparência técnica e impessoal, os algoritmos são criados por seres humanos e operam com base em dados que refletem preconceitos sociais (O'NEIL, 2020).

sobre a inteligência artificial uma arena de reflexões sobre as formas de garantir os direitos humanos neste século.

As consequências desses usos ainda são compreendidas parcialmente, porque é preciso avaliar como as ferramentas de tomada de decisão repercutirão desigualdades, em longo prazo. Ainda assim, os elementos que alimentam essas tecnologias já demonstram um enorme potencial de aprofundamento das desigualdades sociais.

Em uma sociedade em que o uso do ciberespaço está cada vez mais orientado pela captação e pela manipulação de dados, discutir o Racismo Algorítmico[2], como um fenômeno emergente, sob a ótica da análise do biopoder proposta por Michel Foucault, lança luz sobre como os algoritmos podem se tornar instrumentos de controle, ampliar disparidades e exercer poder sobre populações estigmatizadas. Na obra *Infocracia: digitalização e a crise da democracia na filosofia*, Byung-Chul Han (2022) aborda a intensificação da digitalização[3], pelo avanço das tecnologias de informação. Segundo o autor, vivemos hoje em uma sociedade na qual somos controlados e dominados pela constante produção e pelo consumo de informações, sem nos darmos conta disso, em que o poder é exercido por meio da manipulação da informação como instrumento de dominação.

Han (2022) atualiza a visão de Foucault, relativamente ao biopoder e à sociedade disciplinar, para quem o poder é exercido não apenas por intermédio da coerção física, mas também pela vigilância e por práticas de normatização, ao afirmar que, no regime de informação, o controle não se dá mais pela disciplina física, mas sim pela manipulação de informações que influenciam comportamentos. O autor introduz o conceito de “psicopolítica”, que se baseia em como as tecnologias digitais

[2] Ruha Benjamin (2019) define Racismo Algorítmico como um fenômeno no qual as tecnologias digitais e os sistemas algorítmicos reproduzem e reforçam hierarquias e desigualdades raciais, mesmo quando aparentemente projetados para serem neutros ou objetivos. Para Safiya Umoja Noble (2021) refere-se à perpetuação de preconceitos raciais por meio de sistemas algorítmicos que refletem desigualdades sociais e históricas, com impacto em grupos marginalizados. Tarcízio Silva (2022), por sua vez, refere-se à reprodução e à amplificação de preconceitos raciais por meio de sistemas algorítmicos, os quais refletem as desigualdades estruturais e históricas presentes na sociedade brasileira, impactando negros.

[3] Para Han (2022), a definição de digitalização é multidimensional, pois envolve aspectos técnicos, suas implicações sociais, políticas e éticas. A digitalização não se refere apenas à transição de processos analógicos para digitais, é um fenômeno abrangente que modifica significativamente a sociedade e as interações humanas, pois envolve uma mudança estrutural na forma como vivemos, nos comunicamos e nos relacionamos com o mundo e entre nós. Ela está estreitamente ligada ao capitalismo da informação, que opera por meio de coleta, vigilância e controle de dados, alterando nossa percepção da realidade e a eficácia da comunicação.

exploram a psique humana, em que o regime de informação atual exerce controle não mais sobre os corpos, mas sobre a psique, utilizando a coleta e a manipulação de dados para moldar comportamentos. O regime de informação torna os sujeitos “transparentes” por intermédio da coleta de dados, mas, paradoxalmente, essa transparência, apresentada como uma fachada que encobre o controle algorítmico e a vigilância, oculta a dominação.

Dito isso, o objetivo geral desta pesquisa é, justamente, analisar o racismo algorítmico à luz da abordagem do biopoder feita por Foucault. Quanto aos específicos, são os seguintes:

- ✓ Mapear as representações sociais e os estereótipos sobre os grupos racializados; e
- ✓ Compreender como as práticas discursivas que sustentam o racismo algorítmico se relacionam com desigualdades históricas e estruturais.

Para alcançar esses objetivos, realiza-se um estudo teórico-conceitual sobre os conceitos de racismo, viés algorítmico e biopoder, bem como um estudo empírico acerca de alguns casos concretos de racismo algorítmico no Brasil e no mundo, uma vez que estamos diante de um fenômeno em escala global.

Quanto ao recorte para esta pesquisa, delimitou a constituição de um arquivo entre meados do ano de 2022 e o segundo semestre de 2024, coletado em mídias, redes sociais, jornais eletrônicos, disponíveis na internet, por meio do uso da ferramenta denominada Google Alerts[4]. De igual modo, outra fonte de referência foi a *Linha do Tempo do Racismo Algorítmico: casos, dados e reações*, que é parte do projeto Desvelar, de autoria do pesquisador Tarcízio Silva (2023) com um escopo maior, cobrindo casos e dados de danos e discriminação algorítmica, compreendendo o período de janeiro de 22/01/2010 a 26/10/2024, com 176 registros de casos diversos de racismo algorítmico. Essa *Timeline* é parte da pesquisa de doutorado Dados, Algoritmos e Racialização em Plataformas Digitais, desenvolvida pelo autor no PCHS-UFABC. O projeto estuda as cadeias produtivas da plataformização digital (mídias sociais, aplicativos, inteligência artificial), seus vieses e impactos raciais, com casos de racismo algorítmico em suas múltiplas formas no Brasil e no mundo.

[4] O Google Alerts é uma ferramenta que detecta novos conteúdos nas páginas do *Google*, como *posts* para blogs, notícias, artigos e *sites*, e notifica usuários cadastrados via e-mail. Ele permite aos usuários monitorarem a *web* em busca de conteúdo relevante e atualizado, concebido para atender às necessidades de pesquisadores, profissionais de marketing, jornalistas etc. Fonte: https://webcompany.com.br/google-alerts/.

Seguindo esse modelo proposto por Silva, o *corpus* da presente pesquisa se constituiu a partir de exemplos de casos que envolvem: Sistemas de Branqueamento em Filtros de Aplicativos e de Redes Sociais; Reconhecimento Facial Aplicado à Segurança Pública; Sistemas de Recomendação Usados nas Redes Sociais; Testes Padronizados Usados na Saúde e Imagens Geradas por IA (Inteligência Artificial).

No tocante à justificativa da utilização do mecanismo de busca Google Alerts e da adoção da *Timeline* de Silva (2023) sobre vieses e impactos raciais, se estabeleceu por permitir a filtragem de alguns casos de racismo algorítmico que circulam na rede para formar um banco de dados, base para a constituição do *corpus* desta pesquisa, devidamente referenciados e datados no momento da coleta.

A metodologia utilizada nesta pesquisa se baseia em uma abordagem qualitativa e interpretativa, visando compreender os significados e as implicações do racismo algorítmico na sociedade, a partir da utilização de fontes primárias como documentos oficiais dos órgãos criadores dos algoritmos, assim como os desenvolvedores e usuários impactados pelos algoritmos e as observações diretas dos algoritmos em funcionamento. Como fontes secundárias, são utilizados textos acadêmicos, relatórios de organizações não governamentais e artigos jornalísticos sobre o tema.

Nesse campo, esta pesquisa segue alguns caminhos complementares. Assume como caminho metodológico a proposta arqueológica foucaultiana, como base para uma análise do discurso. Para isso, posiciona dois conceitos fundamentais que se interrelacionam no método arqueológico, a saber: o discurso e as formações discursivas. Ao propor uma perspectiva arqueológica para analisar a produção da relação entre saber e verdade, Foucault (2000,2013,1969/2019) enfatiza uma análise que permita compreender como os discursos se formam e como eles produzem conhecimento e verdade. O método arqueológico examina as regras e os princípios que, coletivamente, fazem determinado discurso circular em diferentes épocas. Com base no conceito de formação discursiva, entende-se que há relação entre diferentes discursos que constituem uma formação, se relacionam e influenciam o contexto geral.

Dentro dessa perspectiva metodológica, entende-se como os discursos se transformam e se reconfiguram nos diferentes momentos históricos, mas também como sua existência produz materialidades, desigualdades e relações entre o saber e o poder. Embora a concepção arqueológica de Foucault (1969/2019) esteja presente

de forma mais enfática em *Arqueologia do Saber*, outras obras contribuíram para o que se constitui hoje como uma forma de conduzir uma investigação metodológica sobre o discurso. Em *História da Loucura na Idade Clássica* (1978/2017), assim como em *O Nascimento da Clínica* (1977), Foucault, para analisar a loucura, olha para diversos dispositivos conjuntamente: analisa os discursos dos loucos, dos médicos, dos sacerdotes, dos livros, mas também o funcionamento dos hospitais, como parte da experiência material que, de forma dispersa, garante e atualiza os discursos.

Em *A História da loucura* (1978/2017), Foucault examina o percurso que o discurso sobre a loucura faz, em diversos períodos históricos, e sobre instituições diferentes, principalmente como as narrativas sobre a loucura acompanham as mudanças históricas e sociais sobre o tema. O discurso é o objeto do método arqueológico, porque é possível compreender, a partir dele, o saber e a produção de conhecimento de uma época, assim como examinar o que pode ser dito e quais condições permitem que aquilo seja dito.

Nesse sentido, ordenar o discurso é uma condição para o saber de determinada época. Em *O Nascimento da Clínica* (1977), Foucault analisa a transformação do discurso da ciência médica ao longo dos períodos históricos e como essa mudança reorganiza o saber. Em *As palavras e as coisas* (2000), ele demonstra que, em diferentes épocas, há a formação de uma ordem do saber, ou episteme, que organiza os discursos de maneira institucionalizada.

Esse conjunto de proposições permite considerar que os estudos de Foucault, como fonte de pesquisa, são imprescindíveis para investigar como as instituições e as novas tecnologias digitais exercem controle sobre a vida da sociedade, pela regulação de aspectos biológicos, sociais e políticos. Com base na análise das questões que envolvem o racismo algorítmico, compreende-se de que modo, em certa medida, os algoritmos se tornaram mecanismos de vigilância e governança, fortalecendo e operando discriminações sistêmicas, por meio de formatação e reformulação de comportamentos sociais.

Portanto, este trabalho compreende a dinâmica entre racismo algorítmico e biopoder, para abrir um campo de discussão e explorar o papel que as tecnologias digitais têm no fortalecimento, na propagação e na perpetuação de estruturas de poder na sociedade. Como contribuição, esta pesquisa, dada a amplitude e a complexidade da discussão, do nível de transdisciplinaridade que envolve esses objetos, visa a desenvolver não uma análise propriamente dita sobre do discurso

sobre racismo algorítmico, mas sim desenvolver uma descrição atualizada em torno do racismo algorítmico, a partir de uma perspectiva foucaultiana do biopoder, buscando cooperar com o avanço da discussão sobre o tema. Trata-se de um estudo oportuno, pois o objeto aqui abordado é um fenômeno emergente e complexo, que demanda uma reflexão ética e política sobre o papel dos algoritmos na vida contemporânea de cada um de nós.

Assim sendo, a dimensão da análise do *corpus* nesta pesquisa é sacrificada em prol da consistência e da densidade conceitual e teórica do trabalho em virtude de uma escolha e de um enquadramento do objeto desta pesquisa. O próprio recurso ao enquadramento teórico da analítica do biopoder em Foucault implica esta escolha de não definir um *corpus* sob o qual se vai investir em termos de uma análise mais efetiva.

O interesse pelo estudo do racismo algorítmico à luz da analítica do biopoder do filósofo e historiador Foucault decorre de minha atuação docente, ao longo de uma década, ministrando disciplinas como Análise do Discurso, em turmas de Graduação no Curso de Letras e de Pós-Graduação em Educação para Relações Etnicorraciais, no Instituto Federal de Educação, Ciência e Tecnologia do Pará – IFPA. Por meio da leitura de obras sobre racismo na literatura, discurso antirracista, necropolítica, epistemicídio, entre outros, foi possível observar como o discurso racista está disseminado em nossa sociedade por intermédio dos mais diversificados gêneros e práticas discursivas.

Em 2015, orientei Trabalhos de Conclusão de Curso, na graduação em Letras, voltados para a temática Racismo no livro didático. A partir de então, publiquei artigos os quais tratavam de assuntos correlatos, examinando a materialidade discursiva em textos icônico-verbais, encontrados nos livros didáticos à luz da Análise do Discurso (AD) de linha francesa, fatores linguísticos e ideológicos que permeavam as práticas discursivas nesses textos, com o propósito de compreender em que medida eles reforçam estereótipos cristalizados em nossa cultura sobre determinados papéis do negro na nossa sociedade.

Esse trajeto me conduziu para o estudo do racismo algorítmico, no doutorado, em 2020, o que me direcionou a teóricos que atualizam as discussões sobre o tema, por meio do que eles denominaram de racismo dos dados, racismo algorítmico ou viés algorítmico. Autores como Silva (2019b, 2022a), Noble (2013, 2021), Benjamin (2016,2019), Buolamwini (2017), Gebru (2018), O'Neil (2020), Zuboff (2021), Eubanks

(2018) entre outros, proporcionam um rico campo de teorias como referências sobre o assunto.

O racismo algorítmico como temática de pesquisa científica surge a partir de muitas inquietações sobre como as tecnologias digitais, em especial, o funcionamento dos algoritmos, podem reforçar, consolidar e, em menor escala, combater o racismo em diferentes contextos sociais. É nesse interstício que analisar teorias sobre saber, poder, biopoder e formação discursiva configura-se como salutar para a discussão sobre o tema que emerge no incorpóreo universo digital, tema este que é uma área de investigação emergente e multifacetado, que demanda, conforme exposto ao longo desta argumentação, abordagens interdisciplinares e críticas para compreender como sistemas de inteligência artificial reproduzem e potencializam desigualdades raciais estruturantes.

A complexidade inerente a esses sistemas, somada a obstáculos técnicos, epistemológicos, corporativos e éticos, demanda significativos desafios para pesquisadores, reguladores e para a sociedade como um todo, os quais se deparam com tais problemáticos. Nesse contexto, realizar uma pesquisa que envolve IA, algoritmos e racismo como discurso requer o domínio de aportes teóricos das ciências da computação, da sociologia, do direito, da filosofia, dos estudos críticos raciais, dentre outros (BUOLAMWINI & GEBRU, 2018). Logo, compreender a complexidade técnica e a interdisciplinaridade requerida para analisar sistemas algorítmicos é um primeiro entrave. Essa colaboração interdisciplinar, porém, é frequentemente dificultada por divergências epistemológicas e metodológicas entre as áreas.

Dois agravantes técnicos centrais nesse contexto são a opacidade das “caixas-pretas” algorítmicas – modelos proprietários cujos mecanismos internos são inacessíveis – e a restrição ao acesso de dados e modelos de treinamento, protegidos sob alegação de segredo industrial pelas empresas desenvolvedoras (PASQUALE, 2015). Essa falta de transparência impede a compreensão plena de como os algoritmos operam e tomam decisões. A própria definição e a mensuração do racismo também algorítmico trazem desafios conceituais e metodológicos. Diferentemente de discriminações explícitas, o racismo algorítmico é frequentemente sutil e embutido em vieses estruturais, mimetizados como meros “erros técnicos” (BENJAMIN, 2019). Mensurá-lo quantitativamente – por exemplo, por meio de taxas de disparidade em sistemas de reconhecimento facial – é necessário, mas insuficiente, pois pode falhar em capturar nuances contextuais e sociais profundas.

A origem do problema frequentemente reside nos próprios dados de treinamento: conjuntos massivos que super-representam grupos hegemônicos (como rostos brancos) e sub-representam outros, reproduzindo e amplificando visões estereotipadas e desiguais (BUOLAMWINI & GEBRU, 2018). Empresas tendem a atribuir tais falhas a *bugs* ou imperfeições técnicas, esvaziando o caráter sistêmico da discriminação e evitando sua responsabilização.

Essa dinâmica é intensificada pela opacidade corporativa e pela falta de transparência. Grandes empresas de tecnologia (como Google, Meta e OpenAI) raramente disponibilizam informações cruciais para auditoria independente, incluindo a composição de seus dados de treinamento, os critérios exatos de decisão algorítmica e os resultados de auditorias internas, todos protegidos como segredos comerciais (NOBRE, 2021). Essa postura limita drasticamente a capacidade de pesquisadores externos de identificar, provar e combater vieses discriminatórios de forma efetiva.

Paradoxalmente, vieses também estão presentes no próprio ecossistema de pesquisa acadêmica em IA. A falta de diversidade nos campos de ciência da computação e ciência de dados – com predominância de pesquisadores brancos e do sexo masculino – pode levar à negligência de questões raciais em agendas de investigação (WEST, WHITTAKER & CRAWFORD, 2019). Além disso, o financiamento à pesquisa é frequentemente enviesado, com estudos críticos sobre impactos sociais recebendo menos apoio comparativamente a projetos focados na otimização técnica e na eficiência algorítmica. Conferências científicas de grande porte na área, como NeurIPS e ICML, historicamente, priorizam contribuições de caráter técnico em detrimento de análises sobre consequências sociais, perpetuando uma hierarquia de conhecimento que marginaliza abordagens críticas.

Os desafios éticos e legais são igualmente profundos. Investigações nesse campo precisam equilibrar a necessidade de auditoria algorítmica com o imperativo de proteger a privacidade individual, especialmente ao lidar com dados sensíveis, como aqueles utilizados para classificação racial. A questão da responsabilização também permanece em aberto: na cadeia de desenvolvimento e implantação, é incerto quem deve ser responsabilizado por danos – os desenvolvedores, as empresas ou o Estado (CRAWFORD, 2021). A regulação existente, como o *General Data Protection Regulation* (GDPR) na União Europeia e a Lei Geral de Proteção de Dados (LGPD) no Brasil, embora representem avanços, ainda carecem de

especificidade para tratar de discriminação algorítmica de forma contundente e preventiva.

Por fim, a dinâmica política e a resistência corporativa configuram um obstáculo substancial. Empresas de tecnologia frequentemente exercem forte lobby contra iniciativas regulatórias que imponham transparência algorítmica, como observado em respostas a propostas legislativas nos Estados Unidos e na Europa (ZUBOFF, 2021). Sustentam um discurso de neutralidade tecnológica, que apresenta os algoritmos como entidades objetivas e apartadas do contexto social, ignorando como eles são construídos dentro de estruturas sociais desiguais. Ademais, a militarização e o uso de IA para vigilância governamental – exemplificados pelo policiamento preditivo em bairros majoritariamente negros nos Estados Unidos ou pela vigilância massiva de grupos étnicos, como os uigures na China – ilustram como o racismo algorítmico pode ser instrumentalizado para controle e opressão estatal (BROWNE, 2015).

Em síntese, o combate ao racismo algorítmico demanda superar desafios técnicos, fomentar interdisciplinaridade genuína, romper a opacidade corporativa, diversificar o campo de pesquisa, avançar no marco ético-legal e confrontar as assimetrias de poder que permitem a perpetuação da discriminação sob novas roupagens tecnológicas. Trata-se de um esforço coletivo e urgente para garantir que a inteligência artificial sirva à equidade, e não à injustiça.

Levando-se isso tudo em consideração, este estudo é necessário para compreender as complexas interações entre tecnologia, poder e desigualdade na sociedade contemporânea. E abordar o tema, sob uma perspectiva do biopoder foucaultiano, possibilita explorar como as tecnologias digitais estão entrelaçadas em estruturas de poder, como elas moldam nossas vidas e como podem consolidar, perpetuar ou contestar desigualdades sistêmicas.

Quando à sua estruturação, o trabalho está dividido em sete capítulos. No primeiro, a pesquisa é situada no cenário contemporâneo em torno das novas tecnologias digitais. No segundo, é apresentado o contraponto entre novas tecnologias e antigas questões raciais, as quais perduram o momento presente. Já no terceiro, é apresentado um debate a respeito das teorias que tratam das formações discursivas, item chave para o aprofundamento da tese. Quanto ao quarto, se constitui da exposição acerca de outros conceitos fundamentais para esta investigação, como é o caso de discurso, biopoder, além de uma abordagem do racismo no contexto digital. No quinto, são expostos os conceitos e as manifestações do racismo

algorítmico. No sexto capítulo, é argumentado que o conceito de racismo algorítmico se constitui como uma nova modalidade de formação discursiva, ao passo que no sétimo é apresentado o *corpus* da pesquisa, a partir do qual são desvelados os meandros que permitem afirmar que o racismo algorítmico se faz presente no cotidiano de quem frequenta o ambiente digital e afeta diretamente comunidades racializadas. Após essa sequência, são expostas as conclusões a que esta pesquisa chegou.

1 SITUANDO O ESTUDO

As mudanças trazidas pelo avanço tecnológico não são apenas de ordem prática, uma vez que afetam substancialmente as estruturas de poder e as práticas discursivas. Nesse contexto, à medida que as tecnologias digitais emergem, elas alteram as relações tradicionais de comunicação, criando ambientes para expressão e interação, ao mesmo tempo em que introduzem complexas dinâmicas de controle e vigilância. Elas oferecem meios sem precedentes para a interação entre sujeitos, em escala global, mas são também arenas para manipulação e controle, em que algoritmos coletam e filtram informações de maneira que, muitas vezes, reforçam ideologias dominantes e interesses comerciais. Esse processo, no que tange à discussão sobre o racismo, funda também novas formas de discriminação, associadas aos avanços tecnológicos como braço da dominação racial, mesmo diante de um cenário em que, supostamente, há maior democratização e liberdade.

Esse tipo de debate só se tornou possível devido ao surgimento e ao avanço da internet, a respeito da qual, segundo Silva (2020, p.123), as primeiras formulações teóricas caracterizaram-se por opiniões que consideravam a descorporificação on-line como algo positivo, pela possibilidade de que o corpo social vivesse experiências diversas de sua realidade material. Acreditou-se, nesse período, que se manteria o ambiente digital como sinônimo de virtualidade, de passagem efêmera. Na acepção atual, no entanto, o digital está em todos os lados, e aquela ideia de um *self* diverso, que comportaria múltiplas identidades em cada janela on-line, cedeu à realidade do fenômeno. Havia a compreensão de que o ciberespaço desconsideraria, ou estaria fora, de dinâmicas identitárias, como raça, gênero, nacionalidade e afins.

Tal espaço seria, em tese, uma zona “neutra”. Essa é uma interpretação de um período no qual os ambientes digitais tinham poucas formas de comunicação, com preponderância de uma comunicação textual, e efetivamente era possível “desligar-se do mundo virtual” ao se afastar fisicamente do espaço onde essa virtualidade era acessada. Não havia pesquisadores de populações minoritárias em grande número que se debruçassem sobre o tema, e a suposta neutralidade da plataforma era um discurso comum. Fruto de uma intersecção entre olhares teóricos utópicos e a cegueira racial, não se visualizavam as disparidades raciais em meio digital (SILVA, 2022a, pp.18-19).

Segundo Zuboff (2021, pp.14-15), a realidade digital redefiniu muitas coisas que eram familiares e que eram parte da experiência de um mundo conectado, com consequências parcialmente positivas, porque enriquece a capacidade humana de interação e participação, além de possibilitar muitas perspectivas de contato, evolução e conhecimento. Todos os estratos sociais, e no mundo inteiro, em maior ou menor grau, se relacionam com as tecnologias de informação, o que gera intercorrências também nos dilemas de conhecimento, autoridade, dominação e poder, já que essas categorias estão presentes na forma como a realidade digital se entranhou nas necessidades diárias da vida humana, como mediadora de muitas formas de participação social.

Para autora, essa era, marcada por concentrações de riqueza e de poder relacionados à esfera tecnológica e digital, modifica completamente a ordem coletiva, porque se dissimula como avanço tecnológico e como liberdade nas plataformas. Por trás dessa faceta, no entanto, há uma instrumentalização de dados comportamentais das pessoas, de maneira a incentivar, persuadir e criar comportamentos relacionados ao consumo e ao controle da vida do corpo social. Numa fase anterior, os processos automatizados eram capazes de conhecer comportamentos sociais. Hoje, para além de conhecer, eles são capazes de reconhecer o processo de decisão dos sujeitos e moldar a maneira como essas decisões ocorrem (ZUBOFF, 2021, pp.19-23).

Portanto, pode-se afirmar que o conteúdo apresentado aos usuários em plataformas de redes sociais e serviços de *streaming* baseia-se em complexas análises de comportamento anterior, preferências e interações de rede. Para Saldanha (2023), essa curadoria pode amplificar certas vozes e silenciar outras, criando um campo assimétrico de visibilidade que tem implicações diretas na formação da opinião pública e na concepção de normas sociais. Logo, a tecnologia não é apenas um instrumento ou um meio, ela é uma entidade ativa na modelagem das normas que governam o comportamento social, a comunicação e a identidade.

Ao pensar em tecnologias como essas, que se configuram como fruto da atividade humana, e da subjetividade dos programadores, assim como das escolhas das organizações, é possível questionar quais estruturas racializadas sustentam essas escolhas. Busca-se entender como o racismo algorítmico articula-se nas dinâmicas de biopoder em que o controle de populações e a gestão da vida são atravessados por tecnologias digitais, mas pautadas em diferentes discursos, conforme se pode constatar a partir de proposições de Foucault (1969/2019), em sua

Arqueologia do Saber, em que é apresentada a ideia de que o discurso é um elemento basilar na constituição do saber, do poder, da verdade e do sujeito.

Os discursos só se validam porque há uma organização do saber em torno dessa produção de poder e verdade. A arqueologia como método, que está na base desta pesquisa, se propõe a compreender – não necessariamente decifrar a origem dos discursos – o sistema de dispersão que regula sua formação e sua transformação, objetivando caracterizar não o que se oculta nos discursos, mas as regras a que as práticas obedecem para fazê-los surgir.

Ao tratar o discurso como objeto metodológico, articulam-se especialmente os conceitos aceitos, das opiniões legitimadas, das estratégias de atuação e das condições de produção de uma suposta verdade, implicadas na maneira como as instituições e os sujeitos organizam a ação sobre esse objeto.

Trata-se de uma forma de produção de sentido, o qual é moldado por relações sociais e históricas, sendo responsável por construir a realidade e a subjetividade. Os discursos não são produzidos de maneira descolada, embora sejam dispersos, eles não são apenas teóricos, pois produzem impactos diretos no corpo social, nas práticas, nas instituições e afins. Cada enunciado tem um modo único de existência e produção de sentidos, e, para compreender esses sentidos, é necessário observar quais sujeitos falam nos discursos, a quais instituições eles se filiam e como estão articulados com a história. Foucault argumenta ainda que a verdade é sempre relativa e que não se deve buscar uma verdade absoluta, mas analisar como a verdade é produzida, reproduzida e disseminada por meio de discursos e práticas sociais, e como aqueles são utilizados para exercer o poder e estabelecer relações de dominação (FOUCAULT, 2019).

Sendo assim, os discursos devem ser tratados, em primeira instância, como um conjunto de acontecimentos discursivos, que são cadeias de discursos considerados em determinado tempo específico, e devem sempre ser vistos a partir dos efeitos materiais que produzem. Como tal, eles são influenciados, ao mesmo tempo em que influenciam as práticas sociais, o conhecimento e o exercício do poder (FOUCAULT, 2013). Diversos discursos, associados de maneira pouco homogênea e descontínua, constituem um nexo na produção dos acontecimentos. Há contextos nos quais os surgimentos de discursos são tomados, por meio de diversos procedimentos de controle, para ascender posteriormente como verdade.

Nesse sentido, discursos são submetidos a uma seleção e a um controle, para formar uma regularidade discursiva, no seio de determinado setor ou disciplina. O que isso significa, em linhas gerais, é que as mais diversas áreas da vida humana produzem os seus discursos que, por meio da interação entre os acontecimentos discursivos, formam práticas, ações, perspectivas e verdades (FOUCAULT, 2013).

Já o conceito foucaultiano de formações discursivas é adotado nesta pesquisa como norte metodológico para propor uma abordagem pautada em elementos da análise do discurso. Essa noção é fundamental para compreender como o poder se manifesta na construção de verdades dentro da sociedade. Uma formação discursiva é um conjunto disperso de práticas, regras, instituições e discursos que regulam a maneira como o conhecimento sobre determinada coisa é produzido e disseminado em uma época, com a contribuição de diversas áreas. Ela inclui não apenas a linguagem verbal, em suas manifestações escrita e orais, mas também os gestos, as imagens, as práticas sociais e outros elementos dentro da conformação de determinado conhecimento. Além disso, uma formação discursiva consiste em um conjunto de declarações, conceitos, escolhas e objetos que permitem identificar certa regularidade ou padrão. Nas palavras de Foucault,

> No caso em que se puder descrever, entre um certo número de enunciados, semelhante sistema de dispersão, e no caso em que entre os objetos, os tipos de enunciação, os conceitos, as escolhas temáticas, se puder definir uma regularidade (uma ordem, correlações, posições e funcionamentos, transformações), diremos, por convenção, que se trata de uma formação discursiva (2019, p.47).

Para Foucault (2019, pp.89-90), o discurso não pode ser compreendido apenas em sua continuidade e uma linearidade, porque diferentes produções discursivas atuam na conformação de uma formação discursiva. É, portanto, por meio dos sistemas de dispersão que o discurso adquire posição como verdade, que o dissemina e regula a sua produção. O ponto central desse conceito, para este trabalho, é a formulação de uma perspectiva a qual assume que vários enunciados constituem a prática do racismo algorítmico, porque diferentes práticas e instituições contribuem para sua disseminação. Em meios digitais, entende-se que há um conjunto de representações que, sob a forma de discursos diversos, faz circular os valores racializados por meio dos ambientes on-line.

Em qualquer sociedade, relações de poder estruturam, atravessam e constituem o corpo social, de maneira que é impossível dissociá-las. Certamente,

essas relações sobrevivem por intermédio de diversos mecanismos que unem, traduzem e transmitem poderes, através de instrumentos de referência que pavimentam o poder na sociedade. Elas são sustentadas por uma série de operações que se acumulam, dentre outras coisas, por meio do discurso.

Desse modo, não há exercício do poder que se perpetue sem uma economia discursiva a qual, ao produzir discursos de verdade, repouse com maior tranquilidade. Para Foucault (1998, p.85), nas sociedades modernas, há um elo entre poder, direito e verdade, que se organizam entre si como exigências mútuas. Assim sendo, o poder busca produzir verdade e, ao alcançar esse intento, se assenta por meio de normas e regras do direito. Para entender o poder, e a importância do discurso, é necessário mais do que olhar as formas regulamentares do poder no centro, em foco, nos seus mecanismos gerais. Cabe analisar as brechas, as ramificações, onde o poder – e o discurso – se capilarizam. Nesse sentido, os discursos oficiais podem ser importantes fontes dos mecanismos gerais e dos efeitos permanentes, mas a capilaridade do poder é um objeto central porque mostra como, na prática, materialmente, esse poder é exercido.

Utilizando o conceito de “sistemas de dispersão”, que está intimamente ligado à ideia de formações discursivas, Foucault (2019) aborda uma rede complexa de práticas, instituições, regras e discursos que, juntos, regulam e controlam o discurso sobre determinado objeto. Um discurso racial se erige por meio de um sistema de dispersão, composto por diversas regras, instituições e representações, que operam em contribuição para circular uma verdade e produzir práticas de controle e vigilância a partir dessa verdade institucionalizada. Os sistemas de dispersão são capazes não apenas de transmitir informações, mas de determinar o que é aceitável e o que é verdadeiro e contribuem continuamente para a distribuição de poder dentro da sociedade. Nessa direção, esta pesquisa também se detém nas formações discursivas contidas nas práticas das decisões algorítmicas.

Outra categoria teórico-metodológica relevante neste estudo é o acontecimento discursivo, compreendido como um conjunto de condições mutáveis, que possibilitam a existência de determinado discurso. Não diz respeito apenas às condições discursivas, mas também às não-discursivas, já que há diversas instâncias para compor a autorização institucional de um regime de verdade. Para Foucault, deve-se pensar o acontecimento discursivo como o surgimento de uma nova regularidade, de uma nova ordem, mesmo que compile elementos dispersos. A questão do

acontecimento, para o discurso, é justamente pensar como um determinado enunciado aparece em uma condição, quais elementos sustentam a aparição de um discurso e como essas práticas discursivas produzem efeitos na realidade (FOUCAULT, 2019, p.155).

Neste estudo, abordamos o conceito de “acontecimento” na fase genealógica de Foucault. O acontecimento discursivo e a genealogia do saber estão interligados na medida em que ambos se concentram na análise do discurso e na compreensão das condições de emergência e formação de discursos dentro de contextos específicos de poder. Segundo Foucault (2019, p.35), a noção de acontecimento discursivo é que permite relacionar o acontecimento enunciativo com acontecimentos que são de outra ordem (técnica, econômica, social, política). Na fase genealógica, a ideia de acontecimento é reinterpretada através do conceito de poder. Nesse contexto, o acontecimento não se resume a uma decisão, um tratado, um reino ou uma batalha, mas é visto como uma manifestação das relações de poder que permeiam a sociedade, sendo, segundo Foucault,

> uma relação de forças que se inverte, um poder confiscado, um vocabulário retomado e voltado contra seus utilizadores, uma dominação que se enfraquece, se distende, se envenena e outra que faz sua entrada, mascarada (2021, p.73).

Para Foucault (2019, p.34), o acontecimento não está no campo da repetição ou da inovação, mas no todo que está à volta da produção de um discurso, sobretudo como ele se articula com produções anteriores ou como revela uma condição inovadora em determinado sentido. Ao pensar o acontecimento discursivo como uma dispersão, deve-se refletir sobre como uma cultura, em um dado momento, deixa de fazer certa coisa e passa a realizá-la de outro modo. Então, um acontecimento discursivo é sempre uma descontinuidade, articulada por diversos enunciados em conjunto.

Mais uma das categorias que dão base a esta pesquisa é o conceito foucaultiano de biopoder. Conforme definido por Foucault (1998, 2008) e reiterado por Mbembe (2018b), biopoder é o poder exercido sobre a vida, regulando e controlando os corpos. Está intrinsecamente ligado ao controle da vida e da morte pelo poder soberano, decidindo quem pode viver e quem deve morrer. O biopoder se manifesta na capacidade de regular e administrar a vida das populações, influenciando processos vitais como saúde, natalidade e mortalidade. Ele marca o nascimento da

biopolítica, em que a vida se torna um objeto iminente de poder. O procedimento de “fazer morrer ou deixar viver” muda para um novo paradigma que é “fazer viver e deixar morrer”. No contexto do biopoder, a saúde pública, a higiene, a regulação da natalidade e a gestão da mortalidade são campos primordiais de intervenção estatal.

No tocante à biopolítica, de acordo com Foucault (2021b, 1998), é um conjunto de processos que regulam aspectos vitais, como natalidade, mortalidade e longevidade de uma população. Refere-se a uma expressão do poder sobre a vida, na qual as práticas governamentais permeiam a biologia dos corpos humanos para regular as populações. A ênfase na administração da vida e no controle sobre a morte revela o racismo e a exclusão social como resultados diretos das práticas biopolíticas. A biopolítica introduz instituições assistenciais como mecanismos racionais e seguros, além de outros domínios que se concentram na espécie humana ou em populações específicas em sua totalidade, incluindo seus corpos, a saúde, o ambiente geográfico e as condições sanitárias. Ela estabelece mecanismos disciplinares distintos dos existentes, visando também controlar a população através de previsões, estatísticas e medições globais, utilizadas como dispositivos de análise e controle.

A razão pela qual se adotam, nesta pesquisa, as categorias teórico-metodológicas elencadas anteriormente assenta-se no propósito de se analisar a mecânica, o funcionamento do racismo algorítmico como discurso. Elas funcionam metodologicamente como lentes para construir uma compreensão do objeto em estudo.

Para além da utilização estritamente ligada aos conceitos de Foucault, há outras fontes pertinentes, selecionadas a partir de outro passo de intensa relevância para a presente pesquisa: a revisão bibliográfica sobre o tema em estudo, que permite mapear, reunir e analisar os discursos teóricos acerca do racismo algorítmico, como parte das formações discursivas que explicam o fenômeno. Essa técnica auxilia na compilação de materiais bibliográficos – livros, teses e dissertações – que discutem as representações sociais sobre o negro e as relações raciais no Brasil. Esse *corpus* constitui uma das bases para a pesquisa sobre o racismo algorítmico, na medida em que é possível correlacionar e confrontar a literatura crítica e o discurso racial embutido nas decisões algorítmicas.

2 CONTEXTUALIZANDO O ESTUDO: NOVAS TECNOLOGIAS, ANTIGAS QUESTÕES RACIAIS

Para a exposição desta pesquisa, há que se considerar a necessidade de um traçado histórico em relação a marcos a respeito do avanço tecnológico que permitiu que se chegasse ao momento presente, no qual os algoritmos assumem papel de protagonistas.

Nesse sentido, um primeiro momento a demarcar é 1943, quando o primeiro modelo lógico sequencial de neurônios foi implementado por Warren S. McCulloch e Walter Pitts, dando início ao que se convencionou denominar de inteligência artificial (IA) (MIRA, 2008). Haykin (2008) relata que um sistema de IA pode ser capaz de fazer três coisas: 1) armazenar conhecimento; 2) aplicar o conhecimento armazenado para resolver problemas; e 3) adquirir novos conhecimentos através da experiência.

Essa ferramenta só foi possível também devido ao avanço tecnológico na área dos sistemas computacionais, tanto que Shapiro (1992) destaca que IA é um campo da ciência da computação e da engenharia que se detém na compreensão computacional do que é comumente chamado de comportamento inteligente e com a criação de artefatos que exibem tal comportamento.

Entretanto, qualquer projeto de IA precisa de instruções para funcionar, e essas instruções são justamente o algoritmo, que, segundo Cormen *et al.* (2012), se refere a um conjunto finito de instruções ou regras estruturadas de maneira lógica e sistemática, ordenadas e executáveis, destinadas a solucionar um problema ou a realizar uma tarefa específica, permitindo que um computador ou o sistema processe dados e produza resultados esperados de forma eficiente e precisa. Pode-se pensar em IA como uma tecnologia, não só como um conceito, ou seja, ela é feita através de algoritmos computacionais, que percorrem instruções escritas por meio de uma infinidade de linguagens, que devem ser seguidas pelo computador a fim de que ele execute determinados comandos.

A IA busca criar e aprimorar sistemas computacionais capazes de executar, por conta própria, tarefas que normalmente exigiriam a inteligência humana. Esses algoritmos são ferramentas muito poderosas e que têm fornecido muita ajuda no avanço de diversos campos: da assistência médica (ALAGIĆ *et al.*, 2022; MASTRODICASA *et al.*, 2022; ZHOU *et al.*, 2022), com o notório avanço no diagnóstico precoce de doenças (CARPENTER & HUANG, 2018); no meio ambiente

(TRIVEDI & KHADEM, 2022); no transporte (TSIKTSIRIS *et al.*, 2022); na telecomunicação (HIMANSHU, KHANNA & KUMAR, 2022); na segurança pública (TSIKTSIRIS *et al.*, 2022); nas políticas de governo (ZHU *et al.*, 2022); em várias áreas da computação, mas em especial no reconhecimento facial (BOONIPAT *et al.*, 2022; H, 2019), dentre outras.

Esses usos diversos provam o que afirma Arroyo (2022), quanto a considerar que esse processo envolvendo IA tem se tornado cada vez mais comum para a população de maneira geral, pois, à medida que a população precisa recorrer ao estado ou aos meios eletrônicos, sempre é necessário realizar registros, por meio de dados, seja por biometria, seja pelo reconhecimento facial. Conforme Arroyo,

> Isso ocorre na medida em que as pessoas, em suas relações com o Estado e o mercado, precisam registrar seus dados pessoais eletronicamente ou biometricamente. Isso acontece toda vez que se busca obter um documento de identidade ou passaporte, proceder ao pagamento de impostos, entrar com ação judicial ou quando se registra ao passar pelo controle inteligente de semáforos e suas câmeras (2022, p.140).

Além disso, segundo Elias (2017), a IA é fundamental para se conseguir avanços significativos como o reconhecimento de voz e/ou imagens. Embora isso seja muito eficaz e interessante, o risco de esses algoritmos possuírem algum viés (em inglês, *bias*) é enorme, pois, de modo geral, são caixas-pretas (*black-box*), e em alguns casos são usados para aumentar ainda mais o racismo, como relata Peixoto (2020), em seu estudo sobre os riscos de ampliação de desigualdades quando se trata de temas referente a imigração e refugiados, relacionando a utilização de IA e o direito: “O risco está no uso do algoritmo criando classes de candidatos, acelerando o processo para grupos em função da cor (pessoas brancas seriam encaminhadas para um modo mais veloz)” (PEIXOTO, 2020, p.306).

Dentre os principais algoritmos de IA, destaca-se o aprendizado de máquina (*machine learning*[5]). De modo geral, um algoritmo de IA é um subconjunto estendido de aprendizado de máquina que informa ao computador como aprender a operar por

[5] O *machine learning* (ML) é o subconjunto da inteligência artificial que se concentra na construção de sistemas que aprendem ou melhoram o desempenho, com base nos dados que consomem. A inteligência artificial é um termo amplo que se refere a sistemas ou máquinas que imitam a inteligência humana. O *machine learning* e a IA são frequentemente abordados juntos, e os termos, às vezes, são usados de forma intercambiável, mas não significam a mesma coisa. Uma distinção importante é que, embora todo *machine learning* seja IA, nem toda IA é *machine learning* (https: //www.oracle.com/br/artificial-intelligence/machine-learning/what-is-machine-learning/Acessado em 01/08/2024).

conta própria, ou seja, o aprendizado de máquina é uma forma de conseguir a inteligência artificial, de tal forma que se pode afirmar, com Moacir e Ponti:

> A área de Machine Learning (ML) está interessada em responder como um computador pode “aprender” tarefas específicas como reconhecer caracteres, apoiar o diagnóstico de pessoas com doenças graves, classificar tipos de vinho, separar alguns materiais de acordo com sua qualidade (por exemplo, madeira poderia ser separado de acordo com sua fraqueza, para que possa ser usada posteriormente na construção de lápis ou casas) (2017, p.8).

Bhbosale, Pujari e Multani (2020), a despeito de identificarem pontos negativos na utilização de IA, como a subsituição do trabalho humano por robôs e o seu uso indevido para o cometimento de crimes, asseveram que a principal vantagem da IA é que o trabalho se caracteriza por precisão e maior economia de tempo. Com o crescimento do uso de IA, alguns fatos começam a se tornar evidentes, mostrando o quanto essa tecnologia pode agravar problemas, como relata Silva (2019b) em relação às plataformas digitais, em que se evidencia que o racismo, seja de raça, seja de gênero ou etnia, está diretamente ligado ao nosso contexto:

> os casos de identificação de racismo algorítmico somam-se na medida em que pesquisadoras, ativistas e desenvolvedoras geram relatórios, reportagens e guias de auditoria e ação sobre aspectos discriminatórios em diversos dispositivos midiáticos como análise de recomendação de conteúdo (TUFEKCI, 2015), anúncios (SWEENEY, 2013), reconhecimento facial e visão computacional (BUOLAMWINI 2017; BUOLAMWINI & GEBRU, 2018), buscadores (NOBLE, 2021) e outros. Junto a outros indicativos sobre economia, violência, (necro) política e representação midiática, estes casos lembram que racismo “não deve ser entendido como o comportamento excepcional de indivíduos que se desviam de uma norma social não racista, mas sim como um sistema sociopolítico global" que inclui historicamente formatações dos campos produtivos da tecnologia que favorecem o treinamento enviesado de sistemas que intensificam discriminações e opressões (SILVA, 2019b, p.4).

É evidente que a IA tornou-se permanente e irreversível e está destinada a transformar todo esse processo de atividades repetitivas, análises de grande volume de informações, redução do erro humano, podendo até ser utilizada para evitar que as pessoas corram riscos. Ela pode ser utilizada sem interrupção, estando na base da tomada de decisões de modo rápido, acelerando os procedimentos repetitivos.

Na atualidade, inovações globais em torno da popularização e do crescimento da IA e de novas tecnologias digitais modificaram os modos de vida inteiramente. Essas inovações tecnológicas levaram a profundas mudanças na sociedade, de maneira a fazer com que uma imersão no ambiente digital ocupasse uma esfera quase total na vida de parte significativa da população. Anterior ao universo virtual como um

fenômeno de grandes proporções, houve a entrada gradativa na vida globalizada, a ascensão e a popularização da internet e as alterações nas noções de tempo e espaço, ocasionados pela possibilidade tecnológica de deslocar-se com maior facilidade e de estar conectado a diversos lugares do mundo, por meio de dispositivos, sem precisar sair do lugar.

Esses avanços possibilitaram uma gama de modelos de negócios, comunicação instantânea, interação facilitada, acesso a serviços sem a necessidade da presença física, entre outros. A ascensão dessas tecnologias representa diversos avanços econômicos, sociais e políticos, porque possibilitou uma nova ordem de acesso para os sujeitos e para as organizações, em termos do comércio em plataformas digitais. Mudanças na área da educação, da saúde, da economia e das transações financeiras são alguns dos exemplos do potencial do ambiente digital como um facilitador, que se alastrou de forma que a maioria das necessidades humanas atualmente perpassa pela utilização de dispositivos móveis, seja para pagar uma conta, seja confirmar a presença numa academia, ouvir música no carro, fazer um curso de nível superior, entre outros. Todas essas atividades podem – e são – mediadas em um cenário digital (SILVA, 2021, p.106).

Nesse contexto, cada vez mais, numa interação contínua entre aparelhos e sujeitos, modifica-se a forma de tomar decisões e de se comunicar. As informações que circulam nas redes não são mais alimentadas exclusivamente por humanos, mas por algoritmos formulados por IA que trocam dados, coletam informações e preveem comportamentos em um cenário cada vez mais automatizado.

Essa evolução, no entanto, não ocorre em um ambiente neutro e simétrico, acontece num cenário de profunda assimetria informacional e de poder, em que grandes *players*[6] da tecnologia têm o domínio do modelo de captação e exploração de informações em qualquer dispositivo que esteja subscrito à lógica da economia da informação. Nesse sentido, ecoam as considerações de Foucault e Han relativamente às transformações fundamentais nos mecanismos de poder das sociedades modernas às contemporâneas. Enquanto Foucault (1998, 2020) descreve um poder que atua sobre os corpos e a vida através de instituições disciplinares e reguladoras,

[6] Refere-se aos diversos atores que operam dentro do ecossistema digital, incluindo empresas, consultores, concorrentes e outras partes interessadas. Esses *players* não são apenas as grandes empresas de tecnologia, mas também incluem uma rede mais ampla de entidades que interagem e influenciam o funcionamento do capitalismo de vigilância (ZUBOFF, 2021, p.34).

Han (2025) atualiza essa discussão para a era neoliberal e digital, argumentando que o poder migrou para formas mais sutis de controle, adaptado às novas tecnologias, tornando-se cada vez mais imaterial, porém não menos eficaz.

As inovações, segundo Noble (2021, pp.171-172), são desenvolvidas para coletar incontáveis dados das interações humanas, para além do próprio consumo, mas com referência a ele. Aplicativos de deslocamento, que funcionam por meio de geolocalização, coletam informações sobre o percurso feito pelo sujeito. Essa coleta possibilita a interação entre organizações, para propor sugestões de consumo ao sujeito nessa rota, com base nas parcerias publicitárias extrativistas feitas entre as organizações. A dinâmica é coletar, tratar, analisar e prever, num ciclo por vezes não linear, mas que tem como elemento principal a assimetria de poder dos sujeitos diante da magnitude global das empresas de tecnologia.

À medida que as instituições e a sociedade como um todo ganharam contornos digitais, ou se digitalizaram quase que completamente, com novas plataformas, aplicações e dispositivos, os quais possibilitam uma conexão ativa ininterrupta, novas formas de produzir e consumir informação remodelaram e atualizaram as dinâmicas de ocorrência do racismo. Isso porque o digital alimentou novos tipos de desigualdade racial, por meio, dentre outras variantes, da discriminação algorítmica.

Nakamura & Chow-White (2012, pp.16-18) asseveram que, à medida que a comunicação e as formas de conhecimento mudaram do analógico para o digital, houve uma pressão social sobre a ideia da raça como um aspecto fundamental para a construção da identidade e, sobretudo, como um princípio organizador das relações sociais, como se a economia digital pudesse nublar ou erradicar as desigualdades raciais, que são uma das bases das relações de poder. Não importa, no entanto, quanto “digital” a sociedade tenha se tornado, porque o racismo é um problema contínuo e persiste com as suas atualizações, ao longo das linhas de evolução social.

Pode-se assumir que as tecnologias emergentes consertariam ou bloqueariam os vieses raciais e criariam condições equânimes, porém, muitas vezes, elas reforçam e intensificam o *status quo*. Esse fenômeno foi conceituado no livro *Race After Technology* (*Corrida atrás da tecnologia*), de Benjamin (2019), como “The New *Jim Code*”[7]. É importante refletir que a ideia da neutralidade das tecnologias é construída

[7] Trata-se de uma referência ao Jim Crow, correspondendo a um conjunto de códigos raciais segregacionistas nos Estados Unidos no século XX, que institucionalizou a segregação racial naquele

discursivamente, como um braço do silenciamento dos vieses raciais que estão implicados nas tecnologias. Sob a égide do discurso da neutralidade, há um interesse pela inação, que visa silenciar as denúncias sobre vieses discriminatórios embutidos na concepção e nas decisões nessas tecnologias. Esse discurso é potente porque, embora atualize as práticas raciais, esconde-as em uma falaciosa objetividade.

Portanto, as decisões automatizadas, implicadas nos algoritmos que sustentam as aplicações e os dispositivos, podem reforçar relações de opressão e implementar novas formas de exclusão racial, denominadas por Noble (2021, p.8) como demarcação tecnológica ou *technological redlining*. São um mecanismo de perfilação racial, com inspiração conceitual nas práticas de demarcação urbana de acordo com critérios raciais. A autora explica que raça, gênero e capital se relacionam para produzir condições desiguais e formas de segregação tecnológica. Na arquitetura da economia da informação atual, está embutido o uso onipresente de *softwares*, sem que as pessoas sequer percebam os mecanismos que norteiam a criação desses sistemas.

Seja em sua forma visível, como um resultado estereotipado em uma busca num mecanismo de pesquisa, seja em sua forma invisível, como em uma seleção automatizada de currículos que escolhe apenas pessoas brancas, o cotidiano está atravessado por decisões automatizadas que, acima de tudo, refletem os valores raciais priorizados por esses sistemas.

Em face desse conjunto de considerações, ressalta-se que esta pesquisa se volta para construir uma compreensão sobre questões que tocam a IA e os sistemas automatizados como arquiteturas de controle que expressam valores humanos, pois são formulados de maneiras enviesadas e repercutem essa estrutura racial. No entanto, um dos desafios centrais para compreender como a opressão algorítmica ocorre, sobretudo o racismo algorítmico, é entender que as formulações matemáticas que norteiam as decisões automatizadas são feitas por seres humanos. Ao tratar da evolução tecnológica, muitas vezes, ela é descrita como um fenômeno externo, quase impenetrável à atividade humana, embora produzido por sujeitos e por organizações. Logo, a evolução tecnológica está contida na realidade social e só se perpetua por meio das ações desse corpo social.

---

país. Benjamin (2019) descreve o novo *Jim Code* como o emprego de novas tecnologias que reproduzem desigualdades, mas que são interpretadas e promovidas como objetivas e progressistas, em comparação com as formas de discriminação sistemática existentes anteriormente.

Formulações terminológicas como “big data” e “algoritmos” descrevem a atividade humana aplicada ao desenvolvimento de sistemas, portanto, não são benignas ou neutras. Os grupos que concebem essas arquiteturas informacionais detêm todos os tipos de valores, o que pode promover decisões racistas, sexistas, misóginas e afins. Para Noble (2021, pp.155-156), o sistema, em si, não pode ser racista, mas também não é objetivo, porque a subjetividade está implicada nos atores que o desenvolvem. Algumas vezes, os valores incutidos são abertamente racistas, ao passo que, em outras, os dados que alimentam esses sistemas são enviesados e ensinam ao sistema o mesmo tipo de regra estrutural que rege a sociedade.

Segundo Benjamin (2019), é necessário analisar as tecnologias com certo ceticismo, porque elas escondem discriminações, enquanto aparentam ser mais benevolentes se comparadas com outros recortes históricos. A própria consideração comparativa com arquiteturas históricas de racismo pode ser uma armadilha discursiva. Os desenhos discriminatórios propõem redes de poder específicas em cada época, motivo pelo qual é incipiente analisar em termos de maior ou menor benevolência.

Isso permite considerar que o discurso do progresso tecnológico pode ser facilmente instrumentalizado contra os grupos que mais sofrem com opressões sistemáticas. A diferença das novas arquiteturas raciais é a mudança de uma racialização explícita para um regime de cegueira racial, ou de silenciamento, que destrói as existências de grupos racializados. Nesse contexto, os avanços tecnológicos são vendidos como superiores, porque aparentam estar acima dos vieses, mas eles foram produzidos a partir de uma série de dados que repetem padrões históricos de exclusão e de discriminação (BENJAMIN, 2019).

O’Neil (2020, p.37), por sua vez, faz uma correlação entre o racismo e os modelos preditivos algorítmicos, em que muitos sistemas utilizam apenas uma variável para sustentar e treinar um modelo. Em alguns casos, essa variável única basta, porque o problema a ser resolvido é simples, porém há situações nas quais os modeladores utilizam modelos simplórios e equivocados para tratar de questões complexas. O racismo, em nível individual, é um modelo de previsão enviesado, incompleto, defeituoso. Ele utiliza dados raciais para fazer generalizações. A sociedade é treinada para associar o comportamento de um sujeito como símbolo de sua raça, em uma previsão binária que supõe que todos agem da mesma forma negativa.

Essas suposições raciais vão, muitas vezes, conferir validade para o modelo deturpado dos racistas, e essas suposições tóxicas, transformadas em crenças e programadas para discriminar minorias, ascendem à condição de verdade. A tecnologia também tem impacto através do desenvolvimento de ferramentas de comunicação automatizadas, como *bots*[8] e assistentes virtuais. Essas tecnologias, embora projetadas para facilitar a interação e o acesso à informação, também têm o potencial de distorcer o discurso público, especialmente quando empregadas para disseminar desinformação ou reforçar determinados discursos políticos ou comerciais sem transparência.

A questão é que a tecnologia, ao invés de eliminar esse viés, o camuflou em modelos preditivos, embutiu nos sistemas esse funcionamento cheio de pressupostos prejudiciais, ou, em termos foucaultianos, de formações discursivas racializadas, com consequências diversas. Uma estrutura de poder racializada promove e amplifica divisões sociais, apenas de maneira diferente em cada período. Essa mudança é uma preocupação nesta pesquisa, que considera o racismo algorítmico como uma maneira de amplificar hierarquias sociais e perpetuar os vieses raciais.

Considerados os objetivos deste estudo, interessa de igual modo a esta pesquisa mapear, aos moldes de uma forma arqueológica, a relação entre as práticas discursivas raciais e os mecanismos de produção de vieses racistas nos algoritmos. Propõe-se analisar como os discursos racializados já institucionalizados se espelham nas decisões algorítmicas, assim como há acontecimentos discursivos dispersos nesse fenômeno, que fundam novas formas de discriminação racial.

[8] O termo refere-se a robôs de *software* ou agentes automatizados que executam tarefas específicas em sistemas computacionais. Esses *bots* podem ser programados para realizar ações como responder a mensagens, coletar informações ou automatizar processos. No contexto de viés algorítmico, os *bots* podem ser afetados por preconceitos e discriminação, uma vez que os algoritmos subjacentes podem conter vieses implícitos (SOUSA, 2020).

3 TEORIA DAS FORMAÇÕES DISCURSIVAS

Reconhecido como um dos textos pioneiros da Análise do Discurso de linha francesa, *A semântica e o corte saussuriano: língua, linguagem e discurso* (1971), de Michel Pêcheux, apresenta uma crítica à abordagem estruturalista da linguagem – especialmente à visão de Ferdinand de Saussure, que separa a língua (sistema abstrato) da fala (realização concreta) – e introduz o conceito de formação discursiva para superar as limitações da linguística saussuriana. O autor argumenta que a separação proposta por Saussure ignora o papel do discurso como prática social e ideológica. Pêcheux propõe que o discurso não pode ser entendido apenas como um sistema de signos, mas como uma prática que está ligada a condições históricas e ideológicas.

Roberto Leiser Baronas (2011) assevera que é possível constatar que o gérmen do conceito de formação discursiva aparece no texto *Lexis et metalexis: les problemes des determinants* (1968), de autoria de Pêcheux com Culioli e Fuchs, e seguiu sendo desenvolvido em seus trabalhos posteriores, como no livro *Análise Automática do Discurso* (1969), escrito em colaboração com outros autores. Nessa obra, ele discorre sobre a relação entre discurso e ideologia.

O conceito de formação discursiva de Pêcheux abre diálogo com a obra de Foucault, notadamente com o livro *A Arqueologia do Saber* (1969). Nele, o autor refere-se à formação discursiva para delinear as normas que estruturam os enunciados em uma área específica do conhecimento. No entanto, enquanto Foucault enfatiza as relações de poder e os mecanismos de controle, Pêcheux coloca a ideologia no centro de sua análise, destacando o papel das lutas de classe na produção dos discursos.

Baronas (2011) assevera que aproximar as teorias de Pêcheux e Foucault em relação às noções de formação discursiva e de discurso é problemático, uma vez que as bases epistemológicas dos dois autores seguem caminhos diferentes, com aquele alinhado ao marxismo-leninismo e este a uma tendência historicista. Essas diferenças refletem visões distintas sobre como se deve entender e analisar a relação entre linguagem, discurso e sociedade, revelando as influências filosóficas e políticas que moldam cada um dos pensadores.

A formação discursiva, para Foucault (2019), é entendida como um conjunto de práticas discursivas que se organizam em torno de normas e regras específicas, que

determinam as condições de existência, coexistência, manutenção, modificação e desaparecimento de certos enunciados em um dado campo discursivo, operando dentro de uma história descontínua dos discursos. Essas regras não são explícitas, mas se manifestam nas regularidades de enunciados, pela repetição e pela estabilização de formas discursivas específicas que se estabelecem entre os diferentes elementos que compõem um discurso em diferentes momentos e contextos.

As formações discursivas não se configuram como a soma de enunciados individuais, mas sistemas complexos os quais determinam as condições de possibilidade para a emergência e a legitimação de um dado discurso, compreendido como um sistema de enunciados que estão imersos em uma formação discursiva, em que as regularidades e as descontinuidades coexistem. O autor enfatiza a relação entre discurso e suas condições históricas e sociais, o tratando como um objeto empírico que se desenvolve ao longo do tempo como práticas históricas que determinam o campo do saber, não como simples aglomerado de palavras ou textos, mas como uma prática que obedece a condições específicas de possibilidades.

Assim, as formações discursivas determinam: os objetos de discurso (quais enunciados são possíveis e quais não); as modalidades enunciativas (quem está autorizado a falar); os conceitos (os termos e as categorias utilizados); e as estratégias temáticas (os temas e as teorias que podem ser desenvolvidos, organizando os enunciados em sistemas de poder e saber, influenciando a produção de conhecimento e a construção da realidade social). Por exemplo, a formação discursiva da medicina define quais são os objetos de estudo (o corpo humano, doenças, tratamentos), os tipos de enunciados permitidos (diagnósticos, prescrições, relatórios médicos) e os conceitos utilizados (saúde, doença, cura).

Além disso, as formações discursivas não são neutras, elas estão imbricadas com relações de poder e contribuem para a produção de verdades e normas que podem ser utilizadas para controlar e disciplinar sujeitos e grupos sociais. Trata-se de uma prática que produz efeitos reais no corpo social. Por exemplo, o discurso médico no século XIX não apenas descrevia a doença, mas também criava categorias de “normalidade” e “patologia”, influenciando práticas sociais e políticas.

Segundo Foucault, N’*Arqueologia do Saber*, sobre a definição de prática discursiva,

> Não podemos confundi-la com a operação expressiva pela qual um indivíduo formula uma ideia, um desejo, uma imagem; nem com a atividade racional que pode ser acionada em um sistema de inferência; nem com a "competência" de um sujeito falante, quando constrói frases gramaticais; é um conjunto de regras anônimas, históricas, sempre determinadas no tempo e no espaço, que definiram, em uma dada época e para uma determinada área social, econômica, geográfica ou linguística, as condições de exercício da função enunciativa[9] (2019, pp.143-144).

Foucault argumenta que as formações discursivas são marcadas por descontinuidades e rupturas. Portanto a evolução dos discursos não se dá de modo linear, mas a partir de transformações que refletem mudanças nas estruturas de poder. Além disso, ele destaca que o discurso é rarefeito, controlado por mecanismos de exclusão, como a proibição, a divisão entre o verdadeiro e o falso e os procedimentos institucionais.

Por outro lado, Pêcheux (2010) concebe as formações discursivas a partir de uma base materialista histórica, estreitamente relacionada à luta de classes e à estrutura social, que deve ser analisada em função das condições de produção que a antecedem. Em *Análise Automática do Discurso* (1969), influenciado pelo marxismo e pela psicanálise, ao tratar de formações discursivas, o autor afirma que o discurso é indissociável das formações ideológicas, campo de luta no qual diferentes posições sociais disputam o sentido das palavras e dos enunciados.

Já as formações discursivas são o lugar em que a linguagem e a ideologia se interseccionam e se referem a um conjunto de enunciados que pertencem a uma mesma formação ideológica, determinada pelas posições dos sujeitos no contexto social e pelas relações de classe. Logo, para Pêcheux (2010, pp.163-164), formações discursivas são um sistema de restrições que governa o que pode e o que deve ser dito a partir de uma posição ideológica específica. Numa dada formação discursiva, os sujeitos são interpelados pela ideologia, e o discurso é uma manifestação dessa interpelação, reproduzindo e fortalecendo as relações de poder estabelecidas. As formações discursivas são atravessadas pelo interdiscurso, ou seja, por enunciadas já ditas anteriormente, que influenciam a produção de sentidos no presente. Pêcheux ressalta que os sujeitos não falam de maneira livre, mas dentro das formações discursivas as quais os identificam como sujeitos do discurso.

---

[9] A função enunciativa refere-se ao papel que um enunciado desempenha dentro de um sistema discursivo, incluindo como ele é posicionado e classificado dentro de uma prática discursiva, não se limitando apenas ao conteúdo do enunciado, mas envolvendo também as regras e as condições que determinam sua validade e seu funcionamento dentro de um discurso (FOUCAULT, 2019).

Pêcheux introduz o conceito de interdiscurso, que se refere às relações entre diferentes formações discursivas. Ele argumenta que nenhum discurso existe isoladamente, mas sempre em relação a outros discursos, que podem confirmar, contestar ou transformar seus sentidos. Além disso, enfatiza o papel do esquecimento na produção do discurso: ao falar, o sujeito “esquece” as condições ideológicas que o levaram a produzir determinado enunciado, o que contribui para a naturalização dos sentidos dominantes. O discurso é não somente um conjunto de proposições, mas sim uma construção social que reflete e regula práticas sociais e áreas do saber e deve ser analisada em relação ao conjunto de discursos possíveis, e não como um simples texto ou enunciado empírico.

Desse modo, o discurso é um espaço onde se exercem relações de poder, influenciando e sendo influenciado por elas, seja na forma de controle e regulação, como afirma Foucault, seja na forma de luta ideológica, nas palavras de Pêcheux (2010, 2014). Para eles, os discursos não são neutros, mas permeados por relações de poder e ideologia que moldam nossa percepção da realidade. Por outro lado, para além dessas consonâncias, há outros pontos de afastamento entre algumas posições teóricas. Foucault, por exemplo, adota uma perspectiva arqueológica e genealógica, focando em como os discursos emergem e se transformam ao longo do tempo, sem necessariamente vincular-se a uma estrutura ideológica específica. Ele questiona a noção de um sujeito universal e racional, sugerindo que o sujeito é uma construção das práticas discursivas e das relações de poder. Já Pêcheux, influenciado pela ideologia marxista, vê o discurso como expressão das relações de classe e das formações ideológicas e, embora também veja o sujeito como uma construção, enfatiza como este é interpelado pela ideologia.

Para Foucault, os discursos são como uma “população de acontecimentos dispersos” (2019, p.26). Antes de se ocupar de uma ciência específica, ou de romances, ou de discursos políticos, o material a ser tratado é uma “população de acontecimentos no espaço do discurso em geral” (2019, p.26), e, segundo Foucault, os discursos não devem ser vistos como entidades isoladas ou como um conjunto homogêneo de enunciados. Trata-se de um conjunto finito e limitado de sequências linguísticas que, embora possam ser inumeráveis, estão interligadas por regras e práticas que definem sua construção e seu significado. Essa “população” de discursos é caracterizada por sua dispersão e sua diversidade, o que desafia a ideia de uma narrativa única ou linear na análise histórica.

A análise de tais formações permite um entendimento mais profundo de como os discursos são construídos para perpetuar certas ideologias e como eles são sustentados por práticas institucionais e sociais. No cerne dessa concepção, está a ideia de que discursos são veículos de poder que não somente refletem ou representam realidades sociais e políticas, mas também as moldam. Discursos, portanto, são agentes ativos na formação de subjetividades e na configuração de estruturas sociais e de poder.

Para Foucault (2019), o discurso não se limita à estrutura semântica, ele age como uma máquina que influencia relações de saber, poder e subjetividade. Por meio da enunciação, são criados, produzidos e legitimados objetos sociais, lugares e realidades, agenciando processos de conhecimento, verdades e modos de subjetivação. Foucault (2013) afirma que a criação do discurso é regulada, filtrada, estruturada e redistribuída através de métodos que visam neutralizar seus poderes e riscos, controlar sua natureza imprevisível e evitar sua materialidade densa e ameaçadora (FOUCAULT, 2013, pp.8-9). Logo, os discursos são moldados e regulados por aqueles que possuem o poder para tal, sendo práticas discursivas influenciadas pela posição social do sujeito que os enuncia e pelos contextos sociais em que se insere:

> suponho que, em cada sociedade, a produção do discurso é ao mesmo tempo controlada, selecionada, organizada e redistribuída por certo número de procedimentos que têm por função conjurar seus poderes e perigos, dominar seu acontecimento aleatório, esquivar sua pesada e temível materialidade. Em uma sociedade como a nossa, conhecemos, é certo, procedimentos de exclusão. O mais evidente, o mais familiar também, é a interdição. Sabe-se bem que não se tem o direito de dizer tudo, que não se pode falar de tudo em qualquer circunstância, que qualquer um, enfim, não pode falar de qualquer coisa (FOUCAULT, 2013, pp.8-9).

Para Foucault (2013), as instituições e as forças que as moldam são responsáveis pelo controle e pela regulação dos discursos, estabelecendo limites e orientações sobre o que pode ser dito e quem tem autoridade para falar. Nesse sentido, atente-se para esta ilustração de Foucault:

> E a instituição responde: [...] estamos todos aí para lhe mostrar que o discurso está na ordem das leis; que há muito tempo se cuida de sua aparição que lhe foi preparado um lugar que o honra, mas o desarma; e que, se lhe ocorre ter algum poder, é de nós, só de nós, que ele lhe advém (2013, p.7).

Na dinâmica entre saber e poder, observa-se uma relação interdependente na qual o poder não só molda a obtenção de conhecimento, mas também,

reciprocamente, o saber estrutura e mantém diferentes formas de poder. Os discursos são práticas descontínuas, ou seja, da mesma forma que eles se confluem, podem também se excluir, conforme os interesses de quem os produz, o contexto de produção e o momento histórico a que estão vinculados. Neles, as verdades são construídas e desconstruídas a partir do jogo de interesses e das formações discursivas, que instanciam esses discursos.

Foucault (2013) reitera que os discursos podem ser compreendidos como manifestações de uma verdade que se forma para o sujeito. Esses discursos, ancorados em enunciados concretos, acabam por ganhar um caráter de verdade e passam a ser aceitos socialmente como princípios legítimos e reconhecidos. O poder, nesse contexto, se manifesta de maneira dispersa e não pode ser atribuído a um único grupo ou instituição. Ele circula por meio de redes, práticas e discursos, permeando a sociedade e conferindo-lhe a capacidade de determinar e moldar o que é considerado verdade. Para Foucault (2021),

> o poder não existe. Quero dizer o seguinte: a ideia de que existe, em um determinado lugar, ou emanando de um determinado ponto, algo que é um poder, me parece baseada em uma analítica enganosa e que, em todo caso, não dá conta de um número considerável de fenômenos. Na realidade, o poder é um feixe de relações mais ou menos organizado, mais ou menos piramidalizado, mais ou menos coordenado. [...] Se o objetivo for construir uma teoria do poder, haverá sempre a necessidade de considerá-lo como algo que surgiu em um determinado ponto, em um determinado momento, de que se deverá fazer a gênese e depois a dedução. Mas se o poder na realidade é um feixe aberto, mais ou menos coordenado (e sem dúvida mal coordenado) de relações, então o único problema é munir-se de princípios de Análise que permitam uma analítica das relações do poder (2021, pp.369-370).

O saber, por sua vez, é construído e articulado dentro desses parâmetros de poder, dando origem a verdades que são aceitas e replicadas. Esse processo de construção contribui para a constituição de ordens sociais, pois o que é conhecido e reconhecido como verdadeiro direciona as práticas, as leis e as relações sociais. De acordo com Foucault, o exercício do poder gera novos objetos de conhecimento e acumula informações, já o saber induz a efeitos de poder. Logo, saber e poder não são entidades separadas. Quem detém poder decide o que é considerado conhecimento válido e por quem. Assim, aqueles que produzem conhecimento têm sua visão aceita como verdade, graças a outras formas de poder que possuem, como o político, o acadêmico ou o econômico. Em outros termos, conforme Foucault,

> Temos antes que admitir que o poder produz saber (e não simplesmente favorecendo-o porque o serve ou aplicando-o porque é útil); que poder e saber estão diretamente implicados; que não há relação de poder sem constituição correlata de um campo de saber, nem saber que não suponha e não constitua ao mesmo tempo relações de poder [...] Resumindo, não é a atividade do sujeito de conhecimento que produziria um saber, útil ou arredio ao poder, mas o poder-saber, os processos e as lutas que o atravessam e que o constituem, que determinam as formas e os campos possíveis do conhecimento (2020, p.31).

A tríade de discurso, poder e saber compõe um complexo sistema de relações que sustenta e modula as práticas sociais em diversos contextos. Essa interação desvenda como certos discursos são estabelecidos como dominantes e como estes, por sua vez, moldam as percepções e os comportamentos individuais e coletivos. A investigação desses componentes demonstra que o discurso não se limita a representar a realidade, mas tem o poder de criar essa realidade por meio das relações de poder que ele estabelece e perpetua.

Foucault (2021) assevera que essa relação entre discurso e verdade apresenta um caráter coercitivo:

> a verdade não existe fora do poder ou sem poder. [...] A verdade é deste mundo; ela é produzida nele graças a múltiplas coerções e nele produz efeitos regulamentados de poder. Cada sociedade tem seu regime de verdade, sua "política geral" de verdade: isto é, os tipos de discurso que ela acolhe e faz funcionar como verdadeiros; os mecanismos e as instâncias que permitem distinguir os enunciados verdadeiros dos falsos, a maneira como se sanciona uns e outros; as técnicas e os procedimentos que são valorizados para a obtenção da verdade; o estatuto daqueles que têm o encargo de dizer o que funciona como verdadeiro (2021, pp.51-52).

Assim sendo, o discurso opera como um canal através do qual o poder é exercido, frequentemente camuflando sua própria influência sob a aparência de naturalidade e inevitabilidade. Ao estabelecer limites sobre o que pode ser falado e quem pode falar, o discurso define as condições sob as quais o conhecimento é produzido e legitimado. Segundo Foucault, "as leis são armadilhas: não são, de modo algum, limites de poder, mas instrumentos de poder; não são meios de fazer reinar a justiça, mas meios de fazer servir aos interesses" (2021b, p.90). Essas condições são instauradas não apenas pelos conteúdos explícitos, mas também pelas omissões e pelas exclusões, que desempenham um papel crucial na sustentação das estruturas de poder. De forma complementar, o conhecimento não exerce um papel meramente passivo. Ao contrário, ele intervém ativamente na construção dos discursos e na validação das estruturas de poder.

O saber legitimado em uma sociedade não reflete simplesmente realidades objetivas, mas constitui-se como uma construção que atende a interesses específicos, consolidando posições e normas sociais preestabelecidas. Assim, a produção de conhecimento é simultaneamente produto e processo de práticas discursivas que estão imbricadas em relações de poder. De acordo com Foucault, o poder não reside exclusivamente em uma entidade ou pessoa específica, mas emerge por meio de práticas discursivas, incluindo a linguagem e os conhecimentos disseminados na sociedade. Portanto, nas palavras de Foucault,

> Trata-se […] de captar o poder em suas extremidades, em suas últimas ramificações […] captar o poder nas suas formas e instituições mais regionais e locais, principalmente no ponto em que ultrapassando as regras de direito que o organizam e delimitam, ele se prolonga, penetra em instituições, corporifica-se em técnicas e se mune de instrumentos de intervenção material […] Em outras palavras, captar o poder na extremidade cada vez menos jurídica de seu exercício (2021, p.282).

Para Foucault (2021), esse dispositivo de poder – a maneira como as instituições interagem e como as normas e as práticas sociais são formadas – estabelece aquilo que é considerado verdadeiro, válido. As estruturas de autoridade dão aos seres humanos um senso consciente de suas proibições legais e religiosas quando é permitido dizer determinadas coisas, enquanto o discurso científico e acadêmico cria descobertas e gera fatos, determinando o que pode ser falado. A validade de informações e teorias acrescentadas ao nosso conhecimento do mundo é possível, unicamente, por meio desse dispositivo, que forma e molda a realidade percebida de todos.

Dessa forma, a validade de informações ou teorias é autorizada apenas nesse conjunto de relações de poder, enquanto os discursos são restritos a produtos aceitáveis e aceitos dentro de certos limites. A legitimidade de informações ou teorias, assim, é finalmente encontrada nas relações de poder que produzem a realidade, pois moldam a experiência e a interpretação dela. Esse processo pode ser observado especialmente em campos acadêmicos e midiáticos, em que a definição da validade de informações ou teorias influencia a direção de políticas públicas e das percepções sociais.

Porém, a autoridade do discurso é frequentemente ancorada na credibilidade do conhecimento produzido, que, por sua vez, é um reflexo das estruturas de poder que determinam quais discursos são elevados e quais são marginalizados.

Nesse contexto, Foucault (2013) enfatiza um conjunto de mecanismos relacionados ao controle dos discursos, para quem não se trata apenas de regular o poder contido nos discursos ou enfrentar sua aparição eventual. O foco está em estabelecer as condições para o funcionamento dos discursos e aplicar normas a quem os produz, o que leva a uma restrição no número de sujeitos que podem falar livremente, já que o acesso à ordem do discurso não é irrestrito. Em linhas gerais, essas considerações são sintetizadas por Foucault na seguinte passagem:

> Creio que existe um terceiro grupo de procedimentos que permitem o controle dos discursos. Desta vez, não se trata de dominar os poderes que eles têm, nem de conjurar os acasos de sua aparição; trata-se de determinar as condições de seu funcionamento, de impor aos indivíduos que os pronunciam certo número de regras e assim de não permitir que todo mundo tenha acesso a eles. Rarefação, desta vez, dos sujeitos que falam; ninguém entrará na ordem do discurso se não satisfizer a certas exigências ou se não for, de início, qualificado para fazê-lo. Mais precisamente: nem todas as regiões do discurso são igualmente abertas e penetráveis; algumas são altamente proibidas (diferenciadas e diferenciantes), enquanto outras parecem quase abertas a todos os ventos e postas, sem restrição prévia, à disposição de cada sujeito que fala (2013, pp.34-35).

A partir desse aporte teórico dos estudos arqueológicos e genealógicos do saber, desenvolvido por Foucault, compreende-se nesta pesquisa, como os discursos se organizam, operam e são legitimados em determinados contextos históricos e sociais, notadamente os digitais.

4 DISCURSO, BIOPODER E RACISMO EM SUA FACETA DIGITAL

Este capítulo aprofunda o debate em torno dos conceitos de discurso e biopoder de Foucault como subsídios para esta pesquisa. Examinam-se as condições históricas que possibilitaram o surgimento do discurso sobre racismo algorítmico, investigando as estruturas de conhecimento e poder que o sustentam. Isso inclui a convergência de avanços tecnológicos, a persistência de estruturas racistas na sociedade, a crescente dependência de sistemas automatizados de tomada de decisão e as críticas aos sistemas de poder existentes.

A disseminação da rede global de computadores (*world wide web*) desencadeou transformações significativas na sociedade contemporânea, servindo, dentre outras coisas, como um mecanismo sofisticado de controle e manipulação, moldando percepções e comportamentos de forma sutil e abrangente. Em uma época marcada pelo avanço e pela popularização de novas tecnologias digitais, com mudanças rápidas e profundas na forma como as informações são processadas, disseminadas e controladas, a definição de dispositivos de poder, coerções, regimes de verdade e o conceito de vigilância adquirem novas acepções.

Nesse contexto, as tecnologias digitais não são neutras, elas se consolidam como poderosos dispositivos de formação e regulação de discursos, constituem e são constituídas por eles, moldando a maneira como são percebidas e utilizadas pela sociedade. Esses discursos podem gerar um ambiente de imprescindibilidade relativamente a certos avanços tecnológicos, sedimentando o consenso de que adotar esses avanços é necessário, levando à aceitação passiva de novas tecnologias. Instagram, X, TikTok e Facebook, por exemplo, explicitam essa dinâmica quando modificam a forma de consumo e de disseminação das informações, pois essas ferramentas tecnológicas reorganizam as formas de interação social.

No contexto digital, os algoritmos e bancos de dados não apenas armazenam informações, mas também criam normas que redimensionam a realidade social através de mudanças nas práticas sociais, políticas e tecnológicas, introduzindo novos objetos, conceitos e enunciados, ou modificando os existentes. As plataformas digitais, hoje, constituem um campo no qual este poder-saber se manifesta, configurando o que é visível e o que permanece oculto, o que é dito e o que é silenciado.

Tecnologias de vigilância, por exemplo, são frequentemente justificadas por discursos de segurança e proteção, mas podem ser usadas para reforçar o controle estatal e a repressão. Os valores e os pressupostos incorporados nas tecnologias refletem as perspectivas e os interesses dos seus criadores. Por exemplo, algoritmos desenvolvidos em contextos empresariais podem priorizar a maximização de lucro sobre considerações éticas ou sociais. Esse alinhamento entre discurso corporativo e práticas tecnológicas pode resultar em produtos que amplificam desigualdades e excluem certas populações.

O discurso, segundo Foucault (2019), não se limita a um simples agrupamento de frases ou enunciados, ele se apresenta como uma prática social intricada, que está profundamente interligada ao poder e à geração de conhecimento. Ele não é transparente, pois não reflete uma realidade pré-existente: ele a constrói. É uma forma de poder que pode ser utilizado para dominar, controlar e marginalizar. Assim, as condições de existência do discurso definem quem, sobre quais temas e em quais situações pode falar, e em quais circunstâncias isso ocorre. O discurso, nesse contexto, está sempre permeado por relações de poder, as quais afetam sua criação e sua disseminação.

Foucault (2020) explora a relação entre poder, discurso e práticas punitivas, pois, para ele, o discurso vai além da simples comunicação verbal, ele é composto por práticas e técnicas que influenciam e moldam tanto o comportamento quanto as relações sociais. Ele é usado para legitimar as práticas punitivas, tanto que o discurso jurídico, por exemplo, define o que é crime e como os criminosos devem ser penalizados. O discurso médico, por sua vez, define o que é normal e o que é anormal e pode ser utilizado para controlar o comportamento das pessoas. Também o discurso religioso pode ser utilizado para justificar a desigualdade social. Para o autor, o discurso se articula com outras práticas de poder, como a vigilância e a disciplina, para criar uma sociedade dócil e obediente.

Portanto, as diferentes formas de discurso, conforme Foucault, como exames de si mesmo, interrogatórios, confissões e interpretações, são utilizadas para veicular formas de sujeição e conhecimento:

> É justamente no discurso que vêm a se articular poder e saber. E, por essa mesma razão, deve-se conceber o discurso como uma série de segmentos descontínuos, cuja função tática não é uniforme nem estável. Mais precisamente, não se deve imaginar um mundo do discurso dividido entre o discurso admitido e o discurso excluído, ou entre o discurso dominante e o

> dominado; mas, ao contrário, como uma multiplicidade de elementos discursivos que podem entrar em estratégias diferentes. [...] com o que admite em coisas ditas e ocultas, em enunciações exigidas e interditas; com o que supõe de variantes e de efeitos diferentes segundo quem fala, sua posição de poder, o contexto institucional em que se encontra (1998, pp.94-95).

Em *Verdade e as Formas Jurídicas*, Foucault (2002) apresenta uma definição de discurso que complementa e aprofunda a proposta observada n'*A Arqueologia do Saber*. O discurso, para ele, não é apenas de um conjunto de enunciados, mas um campo de disputa e embate entre diferentes regimes de verdade, cada um com seus próprios critérios de validação e seus próprios mecanismos de poder. A verdade não é algo universal e objetivo, mas sim uma construção social que se produz por intermédio do discurso que não apenas transmite verdade, mas também a produz. Ao mesmo tempo, a verdade que se produz através do discurso influencia e molda as práticas sociais.

Em *A Ordem do Discurso*, Foucault (2013) vê o discurso como uma rede intricada de símbolos, a qual se conecta a outros discursos semelhantes, formando um sistema expansivo que não só documenta, mas também replica e reforça os valores de uma sociedade, garantindo sua continuidade. Assim, o discurso transcende a simples sequência lógica de sentenças e palavras buscando um significado próprio, ele se apresenta como um mecanismo crucial de organização que visa moldar o imaginário social vigente. Nessa perspectiva, o discurso se transforma: deixa de ser meramente um porta-voz de significados disputados ou debatidos para se tornar uma ferramenta de aspiração.

Assim sendo, Foucault (2013, pp.62-63) assevera que o discurso é muito mais que um veículo transparente ou imparcial para desarmar a sexualidade ou apaziguar a política. É, antes, um campo no qual essas forças exercem seus poderes mais intimidadores de maneira destacada. Além disso, o autor condensa as proposições, os fundamentos e as estratégias dessa organização discursiva para, então, explorar as vias de sua análise. Ele apresenta um paradoxo inicial: a complexidade de discutir e questionar o discurso quando se está, inevitavelmente, fazendo uso dele. Isso evidencia a dificuldade de se libertar das táticas discursivas. Sua crítica se volta para os procedimentos discursivos que sustentam e disseminam o controle sobre todas as produções discursivas.

Relativamente aos mecanismos de exclusão externos ao discurso, Foucault (2013, pp.20-21) identifica a interdição, a separação e a vontade de verdade. A

interdição, como mecanismo prevalente, relaciona-se ao tabu sobre o objeto, ao ritual da circunstância e ao privilégio de quem fala, evidenciando a intrínseca relação entre discurso e poder. Por outro lado, a separação ou a rejeição ilustra o distanciamento ou o descarte do discurso emitido pelo sujeito considerado insano, baseado na dicotomia razão *versus* loucura. Por fim, na vontade de verdade, o discurso se estabelece como um meio, uma ferramenta para distinguir o verdadeiro do falso, fundamentado em critérios arbitrários sustentados por instituições e baseados em contingências históricas específicas.

Em torno de outro conceito chave na presente investigação, biopoder, Foucault (2021) refere-se à gestão da vida das populações por meios políticos e sociais. A partir do século XVIII, o foco do poder começa a transitar de uma autoridade soberana, que detém o direito de decidir sobre a morte, para um tipo de poder que busca promover a vida, regulando aspectos diversos da existência humana. O biopoder emerge como uma ferramenta de controle, não mais apenas aplicada ao corpo individual por meio da disciplina, mas estendendo-se à vida da população em seu conjunto. Esse poder se manifesta na capacidade de influenciar processos naturais da vida, como nascimento, morte, produção e doença, visando a otimização da saúde, da longevidade e da capacidade produtiva do corpo social. Assim, o biopoder se insere nas práticas e políticas estatais, marcando o nascimento da biopolítica, em que a vida se torna objeto iminente de poder (FOUCAULT, 2021, p.64).

A partir do debate de Foucault, Mbembe (2018b, p.123) considera que o biopoder está intrinsecamente ligado ao controle da vida e da morte pelo poder soberano, que reside no poder de ditar quem pode viver e quem deve morrer. Nesse contexto, matar ou deixar viver são os limites da soberania, sendo esses os atributos fundamentais do exercício do poder soberano. Essa relação entre biopoder e controle da vida e da morte pelo poder é fundamental para compreender as dinâmicas contemporâneas de subjugação da vida ao poder da morte, conhecida como necropolítica.

Os posicionamentos dos dois estudiosos se complementam e, paradoxalmente, divergem na medida em que focam em um aspecto distinto do par vida-morte: enquanto Foucault privilegia a vida, por meio do conceito de biopolítica, Mbembe direciona seu foco para a morte, com o estabelecimento do conceito de necropolítica.

Foucault argumenta que o biopoder se caracteriza por sua capacidade de “fazer viver e deixar morrer,” invertendo a lógica clássica de soberania que era “fazer morrer ou deixar viver”. Essa mudança paradigmática reflete uma transformação na concepção e na aplicação do poder, evidenciando a transição para sociedades em que o controle sobre a vida se torna um mecanismo central de governança. O biopoder, portanto, opera por meio de uma rede complexa de instituições, práticas e saberes, engajados na regulação dos fenômenos da vida coletiva.

O direito de vida e de morte é basilar sobre a noção clássica de soberania. Trata-se, acima de tudo, do poder do governante de determinar quem morre e quem pode viver. Ele exerce uma autoridade absoluta sobre a vida, manifestando-se de forma impositiva e autoritária, uma vez que tem o direito de tirar a vida, ou seja, de matar. Na esteira do contrato social, os súditos abdicam do poder para garantir o direito à vida, mesmo que este objeto – o direito à vida – devesse estar fora do contrato social, já que ele é a razão de o poder soberano existir. Numa reflexão paradoxal, é justamente o direito de matar que constitui a essência do direito de deixar viver. Foucault inicia essa discussão para teorizar sobre como a raça não desaparece na modernidade, mas se atualiza em um racismo de Estado, que tem ligação direta com a proposição soberana do deixar viver ou fazer morrer, mas a modifica completamente.

Entre os séculos XVII e XVIII, aparecem técnicas de poder centradas no corpo, com procedimentos diversos que organizam a distribuição espacial segundo características biológicas. Esses procedimentos incluem separar, alinhar e vigiar os corpos individuais, mas também populações específicas, o que é tratado posteriormente. Essas técnicas de poder distribuem os corpos espacialmente, mas também os organizam seguindo critérios. Treinar e aperfeiçoar esses corpos também fazia parte de uma economia de poder voltada tanto para vigiar, hierarquizar e inspecionar os corpos, quanto para discipliná-los sistematicamente para o trabalho.

Já no século XVIII, outra modalidade técnica, ou tecnologia de poder, surge para integrar os dispositivos de controle, vigilância e biopoder. Essa não é, per si, uma técnica disciplinar, pois não se dirige ao corpo, mas à espécie. Se as técnicas anteriores estiveram focadas na construção de um modo de individualização, controle e hierarquização, ou seja, do domínio do homem-corpo, a nova tecnologia rege a multiplicidade, a massa global, em processos de nascer, morrer, produzir e afins. Foucault conceitua essa tecnologia como biopolítica e a descreve da seguinte forma:

> Trata-se de um conjunto de processos como a proporção dos nascimentos e dos óbitos, a taxa de reprodução, a fecundidade de uma população etc. São esses processos de natalidade, de mortalidade, de longevidade que, justamente na segunda metade do século XVIII, juntamente com uma porção de problemas econômicos e políticos (os quais não retorno agora), constituíram, acho eu, os primeiros objetos de saber e os primeiros alvos de controle dessa biopolítica (2021b, p.204).

A partir do século XIX, o direito político passou por uma transformação significativa, incorporando uma nova dimensão ao conceito de soberania: a capacidade de promover a vida e permitir a morte. Antes, o poder soberano era caracterizado pela autoridade de tirar a vida ou permitir que ela continuasse. No entanto, com essa mudança, o direito de soberania passou a incluir também a responsabilidade de fomentar a vida, criando uma realidade biológica.

A diferença do nascimento da biopolítica e do biopoder, em relação ao poder soberano anterior, é que se modificam as estratégias de poder. O biopoder e a biopolítica se concentram na gestão da vida [o homem-vivo] e das populações [o homem-espécie]. Decerto, há mecanismos de disciplina dos corpos individuais, já tradicionais, que não perderam lugar a partir do século XVIII, que foram grandes marcas do modelo de disciplina nas mais diversas instituições, como escolas, hospícios, prisões e hospitais. Porém, o biopoder desloca o controle para a regulação em níveis mais amplos. Isso significa cobrir outros aspectos da vida econômica, social e biológica, pois visa não apenas sujeitos em particular.

No entanto, vale ressaltar que muitos problemas sociais, econômicos e políticos compuseram o rol de alvos do controle biopolítico. Ao considerar a observação da população de maneira ampla, as tecnologias de poder servem, nesse momento, para introduzir uma nova forma de agir sobre os fenômenos. As políticas de natalidade, as estatísticas de morbidade, as análises das endemias, tudo isso faz parte de uma série de estratégias de intervenções biopolíticas estatais, introduzindo as instituições assistenciais, como um mecanismo racional e seguro, assim como outros domínios que se debruçam sobre a espécie humana – ou populações particulares – em sua totalidade, os seus corpos, a sua saúde, o meio geográfico em que vivem, as condições sanitárias, entre outros. Essas áreas exemplificam como o biopoder se materializa, influenciando diretamente o bem-estar das populações, com o intuito de otimizar suas condições de vida. Tais intervenções são justificadas por um imperativo de proteção e promoção da saúde, embora também possam servir como pretexto para práticas de exclusão e controle social.

Em *Segurança, território e população*, Foucault (2008) começa a esboçar o estudo de um fenômeno que institui mecanismos biológicos como estratégia de poder. Ele propõe a ideia de que as sociedades modernas reposicionam o fundamento biológico de ser uma espécie. Na medida em que admite que o poder é, em linhas gerais, um conjunto de mecanismos que tem como função manter o *status quo* ou o controle. Esse conjunto de procedimentos é fundamental para entender o desenho desta tese. A primeira consideração relevante diz respeito à centralidade dos mecanismos de poder em todas as relações sociais. O poder não se funda em si mesmo, ele evoca mecanismos como estratégia para estabelecer, manter e modificar as relações, e esses instrumentos são, ao mesmo tempo, o efeito e a causa das relações, pois é possível encontrar o poder em relações visivelmente hierárquicas ou difusas, que geram efeitos encadeados, capazes de sustentar o conjunto dos mecanismos de poder em dado momento, num campo específico. Isso significa que é preciso compreender de que maneira – e quais mecanismos – estão implicados nas relações.

A partir do novo contexto em que a biopolítica se mostra evidente, a população é tomada como um problema de diversas naturezas, seja política, seja científica ou biológica, e apresenta dilemas coletivos, com profundos efeitos econômicos e políticos, mas que acontecem de forma imprevisível. O que a biopolítica faz é fundar mecanismos disciplinares diversos dos já existentes, que visam também controlar a população, com previsões, estatísticas e medições globais, adotadas como dispositivos de análise e controle. A diferença entre essa manifestação do poder e suas formas precedentes, conforme já sugerido anteriormente, é que a disciplina era aplicada sobre o corpo do sujeito, por meio de práticas de vigilância e treinamento. Os considerados loucos, por exemplo, eram separados do corpo social para que a disciplina – assim como outros elementos – pudesse torná-los dóceis. Portanto, o biopoder institui mecanismos globais de equilíbrio, regulamentação e gestão da vida, o que constitui o novo paradigma: o poder de fazer viver e de deixar morrer (FOUCAULT, 2021b, pp.209-210).

É preciso olhar minuciosamente o que significa o fazer viver e o deixar morrer, sobretudo sob o ponto de vista racial, tema central desta pesquisa. Relativamente à vulnerabilidade existencial da vivência do negro, abordada por Frantz Fanon, na obra *Pele negra, máscaras brancas* (2008), o autor traz à tona a visão de que o negro não é visto simplesmente como um homem, mas sim como um “homem negro”,

evidenciando a marca racial que condiciona sua existência. Dessa maneira, o negro ocupa uma zona do não-ser, condenado a ocupar um não-lugar em um universo que, para que o outro seja humano, é preciso que o negro não o seja. Da mesma maneira, é possível ampliar a reflexão para pensar que o biopoder produz um mundo no qual a pessoa negra não é uma pessoa, ela é uma pessoa negra, em uma zona estéril que ocupa outra dimensão dentro do “fazer viver”.

A disciplina compôs a mecânica da soberania, e a morte era a ameaça última para a resistência à disciplina. A morte era um espetáculo público, símbolo do poder soberano assistido por toda a sociedade, assim como uma representação da passagem do poder soberano terreno para o poder soberano divino. Na dinâmica atual, o fazer viver é um ponto sensível, e a morte, em muitos contextos, assume um caráter mais privado. Para fazer viver, o poder controla os acidentes, as eventualidades, a regulação da vida de maneira geral. Algumas populações são suspensas e experienciam uma sobrevida, uma regulamentação precária proposital do biológico. O biopoder centra a dominação na vida, não no corpo, agrupa de forma global diversas tecnologias para controlar e modificar a vida.

Assim, os mecanismos disciplinares atuam diretamente sobre o corpo, moldando o comportamento individual e tratando a anormalidade com punições específicas. Já os mecanismos reguladores operam em um nível mais amplo, intervindo em questões que afetam grupos populacionais, como o controle da reprodução, o acesso a instituições sociais tradicionais e a determinação dos espaços onde certas populações podem viver. Por essa confluência entre a disciplina e a regulamentação, Foucault destaca que há um elemento que circula e une os dois acontecimentos, tanto a disciplina do corpo quanto a multiplicidade biológica. Esse elemento é a norma. Sobre essa concepção, ele reflete:

> A norma é o que pode tanto se aplicar a um corpo que se quer disciplinar quanto a uma população que se quer regulamentar. A sociedade de normalização não é, pois, nessas condições, uma espécie de sociedade disciplinar generalizada cujas instituições disciplinares teriam se alastrado e finalmente recoberto todo o espaço [...]. A sociedade de normalização é uma sociedade em que se cruzam, conforme uma articulação ortogonal, a norma da disciplina e a norma da regulamentação (FOUCAULT, 2021b, p.213).

O biopoder cobre, portanto, toda a superfície e estende o controle do corpo à população com o uso desses dois tipos de tecnologias. A norma é o que assenta o poder sobre o corpo-indivíduo e sobre o corpo-população, um poder que se incumbe de dominar tudo o que há, especialmente as populações consideradas como

indesejáveis. Para Michel Foucault, o poder de regular a vida vai além da simples permissão de viver ou da imposição da morte. Configura-se como a permissão ou não para que determinadas populações vivam, havendo o controle do nascimento dos filhos e a disciplina sobre os corpos para o convívio em sociedade. Trata-se da eleição de populações como inimigas internas e externas, de tentativas de abate simbólico e materiais e de uma sorte de acontecimentos que têm como plano de fundo o biopoder. Nesse sentido, deve-se entender a ideia de fazer viver e deixar morrer de maneira ampla, e o racismo é uma exposição desse paradigma em suas mais diversas facetas.

Categorizar como raça é uma forma de delimitar o que deve viver e o que deve morrer. Toma-se o contínuo da espécie humana e o divide para, ao fragmentar o campo biológico, definir quais grupos são inferiores no interior de uma população. Essa fragmentação do contínuo biológico é um objeto do biopoder. E o racismo de Estado é um mecanismo fundamental da biopolítica porque é a atualização da preservação do poder soberano de matar ou deixar viver. Esse poder permite a uma população extinguir outra, ou parte dela mesma, em nome de sua continuidade. Foucault argumenta que o racismo em uma sociedade de normalização é uma forma de exercício de poder que permite tirar a vida dos outros.

Além disso, o racismo é uma forma de produção de sentidos que é moldada por relações históricas e sociais e que é utilizada para estabelecer relações de dominação e exclusão. Aplica-se sobre um corpo que se quer disciplinar e sobre uma população que se quer regulamentar e tem como objetivo disciplinar, regulamentar e controlar os corpos e as populações consideradas diferentes ou inferiores, como práticas fundamentais para a reprodução das relações de poder e de saber na sociedade.

Outra acepção do racismo, baseada justamente numa ideia biológica, é a de que, para alguém viver, é preciso que o outro morra. O racismo cria a diferença biológica como um problema a ser eliminado, que tem como base a cor da pele e a origem dos sujeitos. A existência de negros é vista como um risco à manutenção da “pureza” racial branca, justificando, sob essa lógica, a necessidade de sua eliminação por quaisquer meios. Para Foucault (2021b, pp.215-216), essa perspectiva é biológica, pois as populações racializadas, tidas como inferiores, são tratadas não como adversários políticos, mas como ameaças à espécie.

Desse modo, o racismo exerce uma função primordial dentro da biopolítica, porque é indispensável, na concepção dos Estados modernos, para decidir quem tem

ou não o direito à vida. É também o que torna aceitável tirar a vida do outro. Nesse sentido, ocupa uma posição *sine qua non* no funcionamento do biopoder dos Estados modernos, o que coloca as populações racializadas pelo poder dominante como indesejáveis em todas as esferas. Acerca da extensão do poder de morte perpetrado pelo racismo, Foucault afirma:

> se o poder de normalização quer exercer o velho direito soberano de matar, ele tem de passar pelo racismo. E se, inversamente, um poder de soberania, ou seja, um poder que tem direito de vida e de morte, quer funcionar com os instrumentos, com os mecanismos, com a tecnologia da normalização, ele também tem de passar pelo racismo. É claro, por tirar a vida não entendo simplesmente o assassínio direto, mas também tudo o que pode ser assassínio indireto: o fato de expor à morte, de multiplicar para alguns o risco de morte ou, pura e simplesmente, a morte política, a expulsão, a rejeição etc. (2021b, p.216).

Quanto às ferramentas de biopoder na era digital, se atualizam por meio do uso de algoritmos, especialmente quando aplicados em contextos como vigilância, seleção de emprego, concessão de crédito e aplicação da lei, funcionando como mecanismos que podem moldar comportamentos, oportunidades e até mesmo a qualidade de vida do corpo social (DA SILVA & ARAÚJO, 2020a, pp.10-11). Assim, os algoritmos tornam-se veículos através dos quais o poder é exercido, não mais apenas sobre o corpo individual, mas sobre a coletividade e suas chances de prosperar dentro de uma estrutura social. Eles são alimentados por valores, normas e preconceitos existentes na sociedade. Essa compreensão é fundamental para desvendar como as tecnologias digitais podem ser empregadas para exercer formas de controle social que refletem e perpetuam desigualdades raciais (DA SILVA & ARAÚJO, 2020a, p.5). A implementação e a aplicação dessas tecnologias refletem uma forma de governo que Foucault descreveria como biopolítica, na qual Estado, corporações e outras entidades exercem um controle detalhado sobre as populações.

Nesse incorpóreo ambiente digital, o poder não é apenas repressivo, mas também produtivo. Ele não apenas proíbe, mas direciona ações, definindo o que é normal e aceitável dentro do espectro social. Desse modo, os algoritmos agem como mediadores do biopoder, influenciando a distribuição de recursos, as oportunidades e os direitos, tal como se pode depreender destas considerações de Tarcízio da Silva:

> Entender o racismo como fenômeno estruturante das sociedades modernas contribui para a compreensão de como as instituições e os processos de subjetivação são moldados na perspectiva do biopoder. As manifestações racistas, a forma como as instituições segregam, a forma como o mercado de

> trabalho seleciona uns em detrimento de outros e mesmo a construção dos estereótipos e fenótipos raciais – cotidianamente reforçados nos meios de comunicação – nos apontam para as formas de funcionamento do racismo (DA SILVA & ARAÚJO, 2020a, p.5).

O racismo é uma expressão de poder baseada na classificação, na vigilância e no controle de grupos étnicos, socialmente construídos e explorados como não-humanos (ALLPORT, 1954; COLLINS, 2019; CORREIA, BRITO, VALA & PÉREZ, 2001; DOVIDIO, 2001; DUCKITT, 1992; FANON, 2008; FERNANDES, 2007; GAERTNER & DOVIDIO, 1986; LIMA & VALA, 2002; MBEMBE, 2018; PETTIGREW, 1958, 1959/1993). Ele é uma estrutura social complexa, arraigada nas instituições, nas relações sociais e nas configurações do poder. Como ideologia, o racismo divide a humanidade em grupos com características físicas comuns, que dão suporte a características morais, intelectuais e estéticas, numa escala desigual. É uma crença, transformada em sistema, de que existem raças naturalmente hierarquizadas, sustentadas pela relação intrínseca entre a aparência física e o intelecto, entre o físico e a cultura.

O racismo cria a raça não num sentido biológico, mas sociológico, porque se pensa não exclusivamente pelos traços físicos, mas pela ideia de que grupos sociais com traços físicos diferentes possuem repertório cultural, linguístico e religioso naturalmente inferior, e tal diferença serve como base para hierarquizar, desumanizar, explorar e justificar a discriminação. Embora a diferenciação de raças humanas não tenha conteúdo do ponto de vista biológico, ela foi a base para a construção de sistemas de dominação em diversos lugares do mundo, de maneira a estabelecer um fosso sócio-histórico impossível de negar, posto que é o racismo – como sistema e estrutura social – que constrói a raça.

Desse modo, pode afirmar que a diferença racial é uma criação humana em contextos de discriminação e subjugação de determinados grupos. Mas, uma vez que a raça foi criada em um contexto sócio-histórico específico, sua existência passou a conduzir a realidade social, mesmo com algumas mudanças ao longo do tempo. A concepção racializada que conduziu a justificativa social para a escravidão de povos africanos nas Américas produziu um sistema em que, mesmo quando formalmente acabado, o regime de escravidão permaneceu como categoria social capaz de validar e perpetuar processos de segregação, desigualdade e dominação.

Esse fenômeno tem repercussões mundiais, ressalvadas as singularidades dos processos históricos em cada região. A raça foi a base para um sistema social

racializado que concedia privilégio sistemático aos europeus – que se tornaram brancos na genealogia do sistema racial – em detrimento de povos não europeus – que se tornaram selvagens, negros, indígenas e não-brancos no dicionário da dinâmica racial.

Esses sistemas sociais se tornaram globais com todos os processos de colonização e escravização que estenderam o seu alcance. Há, portanto, uma estrutura social que, mediante o uso de diversos mecanismos, sustenta o privilégio racial em uma sociedade. A condição de privilégio não é meramente discursiva, embora seja também por meio do discurso que ela se perpetue. Há diversos benefícios materiais, nessa ordem racial, para membros da raça considerada dominante, seja em torno do prestígio, seja da supremacia política ou das vantagens econômicas, o que significa que manter as estruturas sociais hegemônicas é uma missão daqueles que se beneficiam da condição de desigualdade.

Essas estruturas raciais existem e se retroalimentam porque são sustentadas por interesses coletivos daqueles considerados como dominantes. Porém as raças subordinadas lutam para mudar a lógica que opera as desigualdades raciais, em busca de igualdade e equilíbrio, mesmo com o atraso histórico gerado pela estrutura racializada.

O outro braço do racismo é a ideologia racial, que se constitui como uma série de referências utilizadas para explicar, justificar e manter o *status quo* racial. Os quadros de referência não são exclusivos da raça dominante, diga-se de passagem, mas o poder material de determinada categoria de pessoas tende a se tornar, também, o padrão pelo qual outros grupos fundamentam as suas posições. Uma vez que uma determinada parcela dos sujeitos possui o domínio hegemônico dos mais diversos setores da vida humana, imprimem também o seu domínio ideológico, mesmo que parcial.

Essas ideias que justificam e sustentam também o *status quo* racial estão disponíveis especialmente para grupos que, mesmo em condição racial semelhante, fazem parte de outras categorias que se interseccionam na dominação social, como gênero, classe e orientação sexual. Há muitas formas e diversos emaranhados que criam relações de dominação, não apenas a dominação racial, o que explica o porquê de sujeitos do mesmo grupo racial apresentarem interesses múltiplos. Apesar disso, mesmo aqueles que não acessam os privilégios totais por sua situação estrutural específica endossam a ideologia que justifica o *status quo* racial.

Na sociedade brasileira, acreditava-se, após a abolição da escravatura, que os processos de dominação e subjugação dos negros cessariam gradativamente, uma vez que eles estivessem integrados à sociedade de classes. O fim do regime escravocrata, em termos formais, teria deixado populações racializadas em uma condição de difícil e demorada integração à nascente sociedade de classes, além de ter ocupado um lugar social paupérrimo em disputas com os imigrantes europeus que absorveram as melhores oportunidades de trabalho independente, mesmo que em condições subalternas. A sociedade brasileira deixou o negro com a responsabilidade violenta de se transformar para os novos padrões da sociedade de classes, diante do advento do trabalho livre, sem efetivamente integrá-lo economicamente à sociedade e oferecer condições dignas e equitativas de vida (FERNANDES, 2008, pp.210-220).

Uma das análises mais proeminentes no campo das Ciências Sociais, desenvolvida por Florestan Fernandes junto a outras análises de classe da questão racial na Escola de São Paulo, enfatizava a apreciação de um processo de desagregação do sistema escravista e a constituição de uma sociedade de classes. A situação do negro pós-abolição, além de não ter trazido cidadania para ele, permaneceu vinculada ao antigo regime. Dentro dessa análise, compreendia-se que a discriminação racial, o despreparo cultural do ex-escravizado para a nova sociedade e a inadequação do escravizado à posição de trabalhador livre resultaram na subjugação e na desqualificação social do negro para ocupar um lugar na sociedade de classes, embora fosse, para ele, algo temporário. O racismo seria, então, uma reminiscência do passado, modificável após a total integração dos negros à sociedade de classes, em uma formulação ao mesmo tempo otimista e irreal.

Fernandes enfatizava, em sua obra, duas questões centrais: uma avaliação da discriminação racial que viviam os escravizados no mercado de trabalho, inclusive pela preferência por empregar brancos imigrantes europeus; e uma ênfase nas deficiências culturais do ex-escravizado diante do sistema capitalista. Nessa lógica, a discriminação racial seria um elemento da estrutura escravista, mas não possuía compatibilidade com os fundamentos de uma sociedade de classes. O modelo de sistema social da revolução burguesa levaria a um contexto competitivo, com maior potencial para um ambiente democrático e igualitário para os negros. O racismo, como um registro anacrônico do passado escravista, desapareceria com o amadurecimento do capitalismo no Brasil. A conclusão com base nessa concepção é de que o racismo à brasileira seria um fenômeno transitório.

Ramos (1978), por sua vez, ao tratar sobre a discriminação como realidade brasileira, enfatiza que as feridas dessa discriminação são visíveis no olhar mais superficial sobre a realidade vivida. A ideologia oficial consolidada no século XIX sancionava a pública discriminação dos negros no mercado de trabalho, por exemplo, ao permitir de forma consuetudinária expressões como “não se aceitam pessoas de cor”, até meados do século XX.

Mesmo após a proibição categórica em dispositivos jurídicos, a discriminação aspirou ares sutis, mas ainda assim profundamente racistas, como “pessoas de boa aparência”, o que significa, em linhas gerais, uma preferência pela branquidade dos trabalhadores. O racismo pode ser difuso, mas ativo e visível na prática experienciada por grupos racializados. A assimilação social da população negra no Brasil foi uma farsa, no sentido de que, mesmo naqueles lugares onde eram maioria quantitativa da população, como na Bahia, a realidade era a de lidar com e suportar uma enorme discriminação, como minorias econômicas e políticas.

Ramos (1978) faz a denúncia do retrato vivido pelo negro no Brasil em sua época, nos mais diversos setores da vida humana que possibilitam a experiência, ou não, de uma vida digna e plena. Dizia-se, no país, que a condição econômica da população negra não era decorrente da raça, mas há um labirinto de raça e classe que colaborou para esse cenário. A estratificação que reduz as possibilidades de existência de negros persiste não apenas devido à condição de classe desses sujeitos, mas pela sua experiência de raça, num sentido cíclico: vive-se nas favelas por não se ter condições de viver em melhores habitações, e a falta dessa condição econômica é um resultado da discriminação para conseguir emprego. A discriminação no mercado de trabalho tem dois fundos, a falta de preparo e de instrução, mas essa ausência é resultado, também, da falta de recursos financeiros e da dificuldade de acesso à educação. Há, portanto, um ciclo vicioso de discriminação, na sociedade brasileira, que coaduna o trabalho, a escola, a moradia, o acesso à saúde e mais uma série de âmbitos coordenados pela realidade social de que a raça define a posição.

Gonzalez e Hasenbalg (1982), na década de 1980, contribuíram para a crítica ao racismo como uma reminiscência do passado. Apresentaram uma concepção que sustenta a discussão sobre o racismo algorítmico nesta pesquisa: a raça funciona como um critério de distribuição de sujeitos em uma hierarquia racial. A raça é fundamental para a reprodução de classes, ela é a base da experiência da distribuição da estratificação social. O regime escravocrata foi o alicerce das desigualdades raciais

no país, mas o legado hostil e violento da escravidão se torna um ponto mais distante de explicação – total – à medida que se afasta temporalmente do período. Ou seja, a posição social do negro na sociedade contemporânea, como já o era no tempo de escrita dos autores, se relaciona diretamente com as relações estruturais e desiguais entre negros e brancos no presente, não apenas como uma reminiscência do passado.

No que tange à mobilidade social no pós-abolição, e na concepção geral de que o racismo é um resultado de relações estruturais no presente, deve-se considerar que há uma desigual distribuição geográfica de brancos e negros e práticas racistas do grupo racial dominante. Esse último elemento é fundamental porque o racismo gera efeitos para além da dimensão material. Uma organização social racista produz efeitos psíquicos relevantes, que também obstruem a mobilidade e a equidade racial, já que essas práticas discriminatórias fazem com que a população negra internalize autoimagens depreciadas pela ideologia racial (GONZALEZ & HASENBALG, 1982). Essa visão negativa do negro e da negritude aparece em diversos dispositivos, incorporados inclusive nos aparelhos de educação e de saúde, por exemplo, como setores centrais da vida social.

Ainda no campo do trabalho, como empregadores, os poucos pequenos negócios dos negros como vendedores de rua não alcançavam a mesma relevância econômica que têm os brancos empregadores, com o controle dos meios de produção, o conhecimento do mercado e os recursos financeiros para regular e conduzir a economia do Estado. Também na educação, a participação negra é crítica, sobretudo no ensino superior. Destaca-se que Ramos (1978) escreve sobre a sociedade brasileira da década de 1970, mas a realidade, embora com uma significativa mudança de participação negra em diversos setores, segue com a raça, quantitativamente, como determinação da posição social e econômica.

O autor critica a ideia de que os negros foram assimilados e integrados à próspera sociedade de classe, com o exemplo das condições territoriais, na geografia das cidades, a que foram submetidos. O processo de favelização como única possibilidade de morada para os negros, em diversos cantos do país, apresenta um retrato das posições paupérrimas a que tiveram que sobreviver em função das relações raciais no país. Trata-se do nível mais baixo de habitação, que não correspondia às necessidades mínimas de higiene, conforto e saneamento, mas que

restaram como as únicas formas de habitar para uma grande parcela da população, o que configura a segregação habitacional que até hoje se sustenta (RAMOS, 1978).

Nesse contexto, os estereótipos e as representações racializadas reduzem as aspirações dos negros e regulam as possibilidades de sua existência. Esse interdiscurso tem profundo impacto psíquico, mas também interfere na forma e nos limites para que grupos racializados acessem lugares e quebrem barreiras sociais do ponto de vista material. Gonzalez e Hasenbalg concluem:

> Esse perfil de desigualdades raciais não é um simples legado do passado; ele é perpetuado pela estrutura desigual de oportunidades sociais a que brancos e negros estão expostos no presente. Os negros sofrem uma desvantagem competitiva em todas as etapas do processo de mobilidade social individual. Suas possibilidades de escapar às limitações de uma posição social baixa são menores que a dos brancos da mesma origem social, assim como são maiores as dificuldades para manter as posições já conquistadas (1982, pp.98-99).

Bonilla-Silva (2003) ressalta a ideologia racial e como ela é adaptável a diversos contextos, mesmo com a manutenção de sua natureza discriminatória. O racismo pode operar, como é o caso brasileiro, em uma aparência de igualdade racial, uma sutileza social que o mascara sob o mito de uma democracia. É da própria conformação do racismo contemporâneo operar fora do registro escrachado da desigualdade e perpetuar desigualdades estruturais de maneira velada. É mister compreender que o racismo, como o autor relembra, é uma infecção que pode ficar dormente por um tempo, mas não é eliminada, pois muda a forma para se manter. Portanto, o racismo é um legado do passado, mas também uma estrutura maleável que se ajusta às novas configurações sociais, sem perder o componente elementar de sua formação, que é o critério da raça como forma de dominação.

Segundo Fanon (2018), o racismo é o elemento visível e cotidiano de uma estrutura de dominação e é, sem sombra de dúvidas, um elemento cultural, que, como tal, não se torna rígido e precisa se renovar e mudar de fisionomia, para continuar compatível com o todo cultural que o conforma. Também assegura, numa dimensão mais global, a sua compreensão sobre o suposto enrijecimento do racismo, que seria resultado de um mundo com maior equilíbrio social, sobretudo porque as práticas que sustentam a dominação racial deslocaram-se gradativamente da violência física, embora ainda seja uma variável relevante, para manter o sistema em outros meios. A ideologia racial faz surgir uma aparência de democracia racial, pelas violências sutis que tomam o lugar do que antes eram agressões brutais.

Nesse sentido, há que se levar em consideração as seguintes explicações de Fanon:

> À dada altura tinha sido possível acreditar no desaparecimento do racismo. Esta impressão euforizante, à margem do real era simplesmente a consequência da evolução das formas de exploração. Os psicólogos falam então de um preconceito tornado inconsciente. A verdade é que o rigor do sistema torna supérflua a afirmação quotidiana de uma superioridade. A necessidade de apelar em graus diferentes à adesão, à colaboração do autóctone, modifica as relações num sentido menos brutal, mais cambiado, mais "cultivados". Aliás, não é raro ver surgir neste estádio uma ideologia "democrática e humana". O empreendimento comercial de escravização, de destruição cultural, cede progressivamente o passo a uma mistificação verbal (2018, p.83).

Na história dos últimos séculos, o racismo deixou parte da sua feição individual, genotípica e fenotípica, para tomar como objeto uma forma de existir. Como ele é a opressão sistemática de um povo, devem-se pensar os elementos de um povo que oprime, a saber: a destruição dos valores culturais dos povos subjugados, o escárnio às suas modalidades de existência, a desqualificação sistemática de seus repertórios linguísticos, de suas indumentárias e de seus saberes tradicionais em saúde. Ou seja, todo o sistema de referência dos grupos racializados como inferiores são desestruturados, ridicularizados e esmagados.

Não é preciso, no entanto, esforços sobre-humanos para revelar o racismo, nem ele é uma descoberta acidental (FANON, 2018). Embora tome notas de dissimulação, como no caso brasileiro, há muitos elementos visíveis, porque ele se insere num conjunto de explorações de um grupo de pessoas que, ao chegar num desenvolvimento técnico superior, propõe, produz e legitima o racismo, que se estabelece e se atualiza nas sociedades modernas que têm como base o biopoder, especialmente porque o direito a matar é requerido na genealogia desses Estados.

Logo, o racismo é o braço, senão o fundador, da colonização e assume na contemporaneidade a função de sustentar o biopoder. É uma tecnologia do poder, portanto, com funções também biológicas, de purificação da raça, de manutenção do *status quo* racial. Os discursos que significam os negros como indesejados, como criminosos, como inimigos internos e externos perpassam pelo fortalecimento de um discurso biológico, na medida em que o negro, como membro de determinada população, é tomado como uma tábula preenchida pelo conteúdo racial e, portanto, igual a todos os outros que compõem essa população, segundo os estereótipos

raciais. No limite, o racismo retira a humanidade dos negros, como uma das técnicas do “deixar morrer” como prática do biopoder.

Em sentido semelhante, Fanon (2008) explica a construção da humanidade hegemônica como um processo fundado pelo antagonismo. Na dinâmica racial que retira dos negros o direito à humanidade, e à existência plena, sua humanidade é sempre um não-lugar, desautorizado pelas práticas de poder que pavimentam, simbólica e materialmente, uma constante sobrevida. Isso porque, na economia desse poder, o fortalecimento biológico e ontológico de um depende da degradação do outro. O branco não precisa ser reconhecido como humano porque ele é o responsável pelas práticas históricas de poder, por dizer quem é ou não humano. O Estado é um braço desse direito de retirar do outro a vida nas sociedades modernas, da qual um dos fundamentos principais é o não reconhecimento da humanidade de grupos racializados. Ao tratar sobre a perversidade desse reconhecimento, Fanon postula:

> O homem só é humano na medida em que ele quer se impor a um outro homem, a fim de ser reconhecido. Enquanto ele não é efetivamente reconhecido pelo outro, é este outro que permanece o tema de sua ação. É deste outro, do reconhecimento por este outro que dependem seu valor e sua realidade humana. É neste outro que se condensa o sentido de sua vida (2008, p.180).

Segundo o estudioso, para que o branco seja reconhecido em sua inteireza, o negro ocupa um não-lugar e anseia por reconhecimento especialmente porque o reconhecimento de sua humanidade significa uma mudança completa no paradigma do deixar viver e fazer morrer (FANON, 2008). Ao se compreender o racismo a partir do biopoder, é possível constatar como a exclusão sistemática de grupos racializados, em suas formas contemporâneas, segue a mesma lógica de desumanização. É uma exclusão sistemática manifesta, baseada em desigualdades históricas, que se perpetuam e se atualizam.

Na abordagem biopolítica, a noção de um sistema baseado notadamente na gestão da vida e na promoção da existência coloca em evidência a seleção da vida que deve ser preservada e aquela que é dispensável. O racismo, então, é visto como intrínseco à formação do estado-nação moderno, o qual se define por meio de políticas que moldam a população de acordo com determinados critérios étnico-raciais, estabelecendo, assim, uma estruturação racial do Estado. Essa dinâmica de poder, enraizada nos Estados modernos, implica que muitas funções estatais, sob certas circunstâncias, são conduzidas através de práticas racistas.

A concepção de nação, relacionada ao Estado nacional, passa a incorporar a construção de raça e as políticas derivadas dessa construção. A era das grandes navegações no século XV ampliou o contato da Europa com a diversidade humana, mas foi nos séculos XVIII e XIX que a raça ganhou um estatuto científico e se tornou central no pensamento moderno. Sem a intenção de revisitar todos os pensadores que contribuíram para a "invenção das raças", podem-se citar exemplos de figuras que embasaram o racismo em justificativas [pseudo] científicas e antropológicas, promovendo, assim, ideologias nacionalistas, segregacionistas, colonialistas e, eventualmente, políticas eugênicas e genocidas.

Para Mozart Linhares da Silva e Willian Fernandes Araújo (2020a, p.4), esses teóricos compartilhavam a crença na hierarquia racial (entre raças superiores e inferiores) e expressavam aversão à miscigenação, vista por eles como um processo que degeneraria a civilização. Tal abordagem sublinha como o racismo foi e continua sendo um pilar para práticas discriminatórias e de exclusão, influenciando profundamente as políticas sociais e culturais ao longo da história e na contemporaneidade. O racismo científico desempenhou um papel crucial na formação de um discurso nacionalista, entrelaçando-se na definição de "povo" e da "identidade nacional".

Características físicas como cor da pele, estrutura óssea, formato facial, espessura dos lábios e textura capilar foram utilizadas para estabelecer e justificar hierarquias raciais desde o século XIX. Esse paradigma, contudo, enfrentou rejeição significativa após a II Guerra Mundial, com a antropologia e a biologia questionando a validade dos pressupostos defendidos até então. Apesar disso, a ideia de que o racismo científico foi completamente abandonado no pós-guerra não deve ser aceita sem questionamentos, visto que houve esforços para reafirmar a raça como uma categoria científica, mantendo viva sua influência nas sociedades atuais.

O racismo científico interage de várias maneiras com as instituições e com o imaginário social, fornecendo um fundamento epistemológico para uma visão hierárquica das diferenças raciais historicamente estabelecidas. Segundo Da Silva e Araújo (2020a, p.5), o racismo se alinha com a valorização da ciência na modernidade e deve ser visto como um elemento estrutural na organização social contemporânea. O estado-nação recorre ao racismo como parte de sua biopolítica, centrando-se na população para exercer controle e gestão. O racismo, portanto, não é um erro ou um desvio, mas uma forma de racionalidade e uma tecnologia de poder que ganhou

proeminência conforme o conceito de soberania começou a se transformar a partir do século XVII.

Para que o Estado exerça seu poder letal dentro do contexto da biopolítica, que se centra na gestão da vida, ele recorre ao racismo como um mecanismo justificador do extermínio. Nessa lógica, eliminar aqueles considerados nocivos, improdutivos ou simplesmente indesejados é visto como uma maneira de proteger e promover a qualidade da vida da população. Desse modo, o racismo estabelece uma distinção entre aqueles cuja vida é considerada valiosa e aqueles que são vistos como descartáveis. Nesse sentido, tirar uma vida não se limita ao ato de matar diretamente, mas inclui práticas que expõem as pessoas ao risco, como a negligência médica, a exclusão social e a marginalização política. A preocupação não se restringe mais às definições raciais do século XIX, mas também sobre noções de diferença e inferioridade, seja em termos culturais, seja de outras formas de desvalorização da vida.

Atualmente, é crucial reconhecer as novas manifestações do racismo, que podem ser identificadas não só na discriminação direta, mas também em versões culturais, ecológicas, econômicas, institucionais, estruturais, simbólicas e algorítmicas. Compreender o racismo como um elemento fundamental das sociedades modernas ajuda a entender como instituições e processos de formação da subjetividade estão intrinsecamente ligados ao poder biopolítico. As práticas racistas, a segregação institucional, a discriminação no mercado de trabalho e a perpetuação de estereótipos nos meios de comunicação ilustram diferentes dinâmicas do racismo.

Nesta pesquisa, com o propósito de melhor contextualizar a problemática acerca das questões as quais implicam o racismo algorítmico, é necessário discorrer, em linhas gerais, sobre o conceito de branquitude e três categorias de racismo: o subjetivo ou individual, o institucional e o estrutural. O racismo individual, manifesto através de comportamentos e atitudes discriminatórias, é a forma mais visível e, consequentemente, a mais passível de denúncia. Embora mereça condenação, o combate a esse tipo de racismo é limitado se não forem consideradas as estruturas e as instituições que perpetuam a discriminação. Portanto, é essencial voltar a atenção também para as formas institucionais e estruturais do racismo, que moldam as bases da sociedade e influenciam a constituição dos sujeitos sociais.

O conceito de racismo institucional foi inicialmente identificado por ativistas dos Panteras Negras[10], que explicaram como essa forma de racismo nos Estados Unidos transcende as ações racistas individuais, enfocando como as instituições operam de maneira a perpetuar a inferioridade de certos grupos através de mecanismos socialmente invisíveis. Esse tipo de racismo é caracterizado pela sua capacidade de manter grupos marginalizados em situações de desvantagem por meio de processos que não são imediatamente evidentes para a sociedade.

O racismo estrutural, por sua vez, é identificado como uma expressão mais abrangente do racismo, influenciando profundamente a estrutura social. Ele se manifesta nas divisões de classe, nos estereótipos étnico-raciais, nos hábitos e nas linguagens, organizando o sistema de privilégios de forma a penetrar em diversos aspectos da vida social e cultural, bem como nas percepções inconscientes.

No início do século XX, intelectuais como W. E. B. Du Bois (1935), Frantz Fanon (*Pele negra, máscaras brancas*, 1952), Alberto Guerreiro Ramos (*O problema do negro brasileiro*, 1957) e, mais recentemente, Maria Aparecida Bento (*Psicologia Social do Racismo*, 2009), entre outros, apresentam as primeiras definições de branquitude. Diretamente relacionada ao racismo, é impossível discutir um sem abordar o outro. Ela se baseia na falsa concepção de uma superioridade racial branca. Em sociedades racistas, esse conceito leva a uma condição de privilégio, em que sujeitos classificados como “brancos” possuem vantagens simbólicas e materiais à custa dos “não brancos”. Essas práticas têm estabelecido a supremacia branca como uma estrutura global, afetando amplamente não apenas relações econômicas e políticas, mas também construções sociais e de identidade.

Porém, a construção da branquitude como um ideal de humanidade é um fenômeno que transcende a mera pigmentação da pele, inserindo-se num contexto mais vasto de normas culturais, sociais e estéticas. Historicamente, a branquitude foi erigida como sinônimo de civilidade, inteligência e beleza, parâmetros que foram

[10] Os Panteras Negras foram um movimento político e social, surgido na década de 1960, nos Estados Unidos, durante o auge dos movimentos pelos direitos civis. Fundado por Huey P. Newton e Bobby Seale, em Oakland, Califórnia, o Partido foi uma resposta direta à brutalidade policial, à segregação racial e à opressão sistêmica enfrentada pela comunidade afro-americana. O movimento adotava uma abordagem mais radical, defendendo a autodefesa armada e a autodeterminação da comunidade negra. Adotaram uma ideologia revolucionária, influenciada por pensadores como Karl Marx e Frantz Fanon. Seu programa incluía demandas como a autonomia para a comunidade negra, o direito a emprego pleno, moradia digna e educação que ensinasse a verdadeira história dos negros nos Estados. Fonte: Copyright © 2025 História do Mundo. Tema Astra para WordPress, em https://nossahistoria.net/pantera-negra-o-movimento-e-a-luta-pela-autodeterminacao/.

utilizados para justificar a dominação e a exploração de povos considerados “outros”. Essa idealização da brancura como norma tem raízes profundas na história ocidental, sendo perpetuada por meio de instituições, práticas culturais e políticas.

De acordo com Drumond (2021, pp.28-29), esse processo de normalização da branquitude como padrão de humanidade criou um sistema de privilégios que beneficia sujeitos brancos em diversas esferas da vida social. Na esfera econômica, por exemplo, esse privilégio se manifesta através de melhores oportunidades de emprego, remuneração mais elevada e maior facilidade de acesso ao crédito. No campo educacional, observa-se uma tendência de instituições de ensino de maior prestígio possuírem um corpo estudantil majoritariamente branco, refletindo desigualdades no acesso e na qualidade da educação.

A branquitude cria filtros de percepção da competência e da inteligência que contribuem para a manutenção de estereótipos racistas que descreditam saberes e habilidades de sujeitos não brancas. Essas barreiras invisíveis impostas por uma minoria privilegiada limitam a ascensão profissional do sujeito racializado, reiterando ciclos de desigualdade e exclusão social. A normalização da branquitude revela-se na segregação espacial, na representação desproporcional em posições de poder e na visibilidade midiática. A desigualdade racial é justificada pela permanência desse sistema de privilégios por meio de discursos meritocráticos, que minimizam os fatores estruturais que desfavorecem grupos racializados. A meritocracia é um discurso ideológico que torna invisível as barreiras sistêmicas que grupos raciais minoritários enfrentam e, assim, oculta as vantagens estruturais àqueles que a sociedade considera “melhores”, ou seja, sujeitos brancos. Dessa forma, é atribuído às minorias o fracasso étnico a prováveis limitações individuais ou culturais, em vez de considerar a operação de um racismo estrutural.

Essa noção, de fato, ainda tem inúmeros impactos no mundo real, refletida na abordagem de muitos profissionais de saúde para com os pacientes negros e suas necessidades. Em última análise, a ideia de que os negros são capazes de sentir menos dor que seus outros pares brancos é uma singularidade profundamente enraizada do racismo estrutural que pode ser rastreada até os períodos colonial e escravocrata. A crença no “mito da menor dor” de seres humanos negros foi uma das bases da construção de discursos que lhes negavam a humanidade intrínseca e, portanto, a dor.

Infelizmente, esse conceito permeia o campo da medicina até hoje, afetando o atendimento prestado aos pacientes negros pelos sistemas de saúde. Como resultado desses vieses envolvendo a raça, os profissionais de saúde frequentemente subestimam a dor sentida pelos pacientes negros ou são menos inclinados a administrar o remédio opioide necessário para aliviar o sofrimento desses sujeitos, em comparação com o tratamento oferecido aos sujeitos brancos. Essa diferença na prestação de cuidados leva a defasagens substanciais na qualidade de atendimento médico prestado, o que equivale à injustiça e ao sofrimento desnecessários.

A origem desse estereótipo está intrinsecamente relacionada à justificação da violência física e psicológica anteriormente cometida durante o período escravocrata. No século XVIII, a crença na resistência dos negros à dor era utilizada como um forte argumento para punição brutal corporal e para o trabalho duro. Essa abordagem serve não apenas para humilhar e intimidar, mas, acima de tudo, mantém a ordem social hierárquica, permitindo que os colonos-brancos e senhores preservem o monopólio do poder sobre seus escravizados.

Hoje, a manutenção desse estereótipo indica sua persistência em níveis mais altos do que a mera prática do aconselhamento. Esse achado se correlaciona não apenas com questões de saúde, violência e interação, mas também em termos de policiamento, estado de direito e decisões judiciais, em que grupos racializados, de modo sistemático, enfrentam uma maior desumanização, estando sujeitos à ação policial mais brutal, com sentenças judiciais mais severas, em um ciclo vicioso de estigmatização e marginalização.

Para Da Silva e Araújo (2020a, p.6), o racismo influencia a formação da identidade do corpo social, participando ativamente na sua constituição. Investigar os mecanismos que promovem essa dinâmica é crucial para compreender como o racismo se estabelece como uma norma socialmente aceita. Isso inclui o exame de como discursos veiculados por meios de comunicação, literatura, arte e outros elementos culturais contribuem para essa normalização. Esta pesquisa centra-se particularmente no que se considera ser o mecanismo contemporâneo mais significativo nesse processo: a política, ou mais precisamente, a biopolítica que emprega algoritmos.

Para Benjamin (2019), o racismo algorítmico é uma forma de biopolítica digital, na qual os corpos racializados são regulados, normatizados e controlados por sistemas automatizados que perpetuam desigualdades estruturais sob o disfarce de

neutralidade tecnológica. Desse modo, essa forma de controle algorítmico pode ser entendida como uma nova expressão da biopolítica foucaultiana, em que o poder se exerce por intermédio de mecanismos que parecem objetivos e eficientes, mas que, na verdade, estão profundamente enraizados em hierarquias raciais.

As tecnologias, porém, não criam o racismo, mas o modificam e amplificam, porque a mediação dessas tecnologias está permeada pelos ideários raciais de desumanização da população negra. É um fenômeno atrelado à desumanização de grupos racializados, que coaduna práticas tradicionais de exclusão social – o racismo – e novos dispositivos e campos de poder. Nesse sentido, a tecnologia não apenas revela a discriminação, ela também produz a discriminação ampliada.

5 RACISMO ALGORÍTMICO: CONCEITOS E MANIFESTAÇÕES

Nesta investigação sobre racismo algorítmico, é essencial analisar casos concretos, com base em ocorrências reais de manifestação desse fenômeno, nos diversos setores da sociedade, para compreender como essas ferramentas digitais podem reconfigurar, atualizar e perpetuar casos de discriminações raciais.

O desenvolvimento de algoritmos e de sistemas de inteligência artificial (IA) desempenhou um papel central nesse processo, ao ponto de algoritmos serem agora fundamentais para decisões que afetam desde o acesso a serviços essenciais até a definição de políticas de segurança pública. No entanto, apesar da sustentação de um discurso de neutralidade e eficiência, esses sistemas têm reproduzido, redesenhado e, em alguns casos, exacerbado desigualdades sociais e discriminações preexistentes, fenômeno que estudiosos têm denominado de “racismo algorítmico” (NOBLE, 2021).

O avanço da tecnologia digital (NOBLE, 2021; BENJAMIN, 2019), nas últimas décadas, impactou de maneira significativa as dinâmicas sociais, políticas e econômicas. Dessa forma, os algoritmos têm se tornado ferramentas essenciais em uma variedade de contextos, desde o marketing até a aplicação da lei, prometendo eficiência e neutralidade. Os algoritmos de busca, por exemplo, podem perpetuar estereótipos raciais prejudiciais, com resultados que, frequentemente, reforçam visões negativas acerca de minorias raciais. Esse fenômeno não se restringe apenas a ocorrências observadas em sistemas de busca na internet, os sistemas de inteligência artificial, usados em setores como o judiciário, a educação e a saúde, têm sido criticados por reproduzirem preconceitos de forma automatizada.

Para entender o racismo algorítmico no contexto das estruturas de poder contemporâneas, é necessário analisar esse fenômeno à luz do conceito de formação discursiva. Não se trata apenas de uma consequência do avanço tecnologia, mas de uma representação das desigualdades sociais e raciais que são mediadas e perpetuadas por sistemas automatizados, uma vez que os algoritmos não se configuram como ferramentas imparciais e podem, de modo sistemático, refletir e amplificar os preconceitos e as desigualdades raciais observados na sociedade. São decisões automatizadas, as quais operam em situações da vida cotidiana, como nos processos seletivos de candidatos a vagas de emprego e nas

decisões na esfera judicial, as quais carregam os vieses dos dados usados para treiná-los.

Em face desse contexto, os algoritmos podem prejudicar, de maneira sistêmica, determinados grupos sub-representados. Os danos não são causados unicamente por dados distorcidos e pela história, mas também pela maneira como os algoritmos são usados e interpretados em determinados contexto. Um exemplo disso é o sistema de reconhecimento facial, que inclui principalmente fotos de sujeitos brancos. O sistema, por esse motivo, é menos preciso na identificação de negros, o que, por sua vez, gera prejuízos e sequelas notadamente para grupos racializados. Um sistema bancário pode levar isso em consideração muito mais indiretamente, por exemplo, avaliando os bairros ou os critérios de emprego, gerando, com isso, desvantagens econômicas. Isso demonstra como os algoritmos podem legitimar disparidades existentes, disfarçadas de objetividade.

Silva (2020, pp.131-132) postula que o entendimento do racismo algorítmico exige uma análise crítica das etapas de desenvolvimento e aplicação de algoritmos, incluindo a coleta de dados, o design do modelo e a interpretação dos resultados. Cada uma dessas etapas pode introduzir ou amplificar vieses, dependendo de como os dados são selecionados, como os algoritmos são codificados e como suas saídas são aplicadas. Assim, a problemática do racismo algorítmico é intrinsecamente ligada a práticas e escolhas feitas por cientistas de dados, engenheiros e tomadores de decisão, que podem, consciente ou inconscientemente, codificar suas próprias suposições e seus próprios preconceitos nos sistemas que criam.

A discursividade algorítmica, entendida como a maneira pela qual os algoritmos efetuam a mediação, a reprodução e a disseminação de discursos, configura-se como um terreno fecundo para a análise das dinâmicas contemporâneas de poder e identidade. Por intermédio de seleção, categorização e exclusão de informações, os algoritmos desempenham um papel central na construção da “outridade”, processo pelo qual determinados grupos são marcados como diferentes, marginalizados ou até mesmo desumanizados. Essa capacidade de os algoritmos influenciarem a construção discursiva reflete e amplifica as relações de poder existentes na sociedade, contribuindo para a manutenção de hierarquias sociais e para a perpetuação de desigualdades.

Desse modo, os algoritmos, operando como agentes não neutros na mediação da informação, segundo Silva (2020, p.176), moldam ativamente os discursos ao

privilegiar certas vozes e perspectivas em detrimento de outras. Ao fazer isso, eles reforçam a visibilidade de alguns grupos enquanto silenciam ou marginalizam outros, efetivamente participando na construção da “outridade”. Esse processo não é meramente incidental, mas resultante de decisões de design e implementação que refletem valores, preconceitos e objetivos específicos dos seus criadores, bem como das instituições que os empregam.

A categorização algorítmica[11], elemento central na organização da informação digital, frequentemente recorre a simplificações e generalizações que podem distorcer representações sociais. Essas práticas correm o risco de cristalizar estereótipos e reforçar divisões sociais, ao invés de promover uma compreensão mais rica e multifacetada das identidades. A “outridade” construída algoritmicamente, portanto, não é apenas um reflexo de diferenças existentes, mas uma ativa contribuição para a estruturação de formas de ver e entender o mundo que são profundamente marcadas por relações de poder assimétricas.

A exclusão de informações, seja por filtragem, seja pela marginalização nos sistemas de recomendação, atua como um mecanismo poderoso para a manutenção da “outridade”. Ao limitar o acesso a perspectivas e experiências diversificadas, os algoritmos promovem uma homogeneização do discurso que favorece o *status quo*, dificultando o questionamento de discursos dominantes e a emergência de contradiscursos.

Essa dinâmica algorítmica de construção da “outridade” e de reprodução de discursos de poder não ocorre no vazio, ela interage com práticas sociais mais amplas. Desse modo, os algoritmos, ao serem empregados em plataformas de mídia social, sistemas de busca e outras tecnologias de informação, inserem-se em um ecossistema de comunicação que é intrinsecamente ligado às estruturas de poder existentes na sociedade. Assim, eles atuam não apenas como espelhos dessas estruturas, mas também como agentes que as reconfiguram e as reforçam.

A discussão sobre racismo algorítmico também implica reconhecer a opacidade que muitas vezes envolve os sistemas algorítmicos, já que há falta de transparência

[11] Eubanks (2018) define a categorização algorítmica como o processo de usar algoritmos para classificar e organizar pessoas em categorias predefinidas. Ela argumenta que essa prática, embora pareça neutra e objetiva, pode ter consequências negativas para grupos marginalizados, como o reforço de estereótipos negativos e preconceitos sobre grupos minoritários e a discriminação algorítmica pela discriminação de grupos minoritários no acesso a serviços públicos, como moradia, saúde e educação.

sobre como os algoritmos tomam decisões específicas, dificultando, desse modo, a identificação e a correção de vieses. Essa opacidade é agravada pelo uso de algoritmos de aprendizado de máquina (*machine learning*) e IA, cujos processos de tomada de decisão são notadamente complexos e inacessíveis para aqueles fora do campo da ciência da computação, incluindo os que são impactados por essas decisões automatizadas.

A sobredeterminação algorítmica, característica marcante da sociedade contemporânea, emerge como um mecanismo de poder que transcende a mera observação, inserindo-se profundamente nas dinâmicas de normalização social. Essa forma de vigilância, mediada por algoritmos, não apenas monitora, mas também classifica e molda comportamentos e identidades. Através de coleta e análise de dados em larga escala, os algoritmos têm o poder de definir o que é considerado normal, desejável ou desviante, influenciando, assim, a percepção social e individual. A normalização, nesse contexto, não se limita a uma questão de conformidade com normas sociais amplamente aceitas, ela se estende à imposição de padrões específicos que, muitas vezes, refletem e reforçam preconceitos e estereótipos existentes. Um dos aspectos mais preocupantes dessa dinâmica é o reforço de estereótipos raciais, em que algoritmos podem amplificar e até perpetuar discriminações, ao basearem suas análises em dados históricos[12] e sociais já enviesados.

Portanto, é inegável que os algoritmos, ao operarem com base em padrões pré-estabelecidos, por vezes, excluem ou marginalizam grupos sociais, especialmente aqueles já impactados por desigualdades estruturais, como é o caso de diversas comunidades racializadas. Assim, o enviesamento algorítmico não é apenas uma questão técnica, mas também profundamente política e ética, refletindo as desigualdades e os preconceitos da sociedade em que essas tecnologias são desenvolvidas e aplicadas.

[12] Referem-se a informações e estatísticas que documentam a presença e os impactos do racismo ao longo da história do país, como a escravidão dos negros trazidos da África, que, no Brasil, durou mais de três séculos, e a desigualdade racial, persistente no Brasil. A taxa de informalidade entre a população branca era de 32%, enquanto entre os pretos e pardos era de 43% e 47%, respectivamente, no ano de 2021, segundo o IBGE (2018). Essa desigualdade também se reflete na educação e no mercado de trabalho (REZENDE, 2025).

A falta de transparência e de *accountability*[13] nas operações algorítmicas apenas agrava esse problema, limitando a capacidade de sujeitos e comunidades de contestar ou mesmo entender as decisões que os afetam. Nesse cenário, a vigilância algorítmica atua como um mecanismo de controle social que se estende para além da esfera pública, invadindo a privacidade e a autonomia individuais. A normalização imposta por esses sistemas não apenas limita as possibilidades de expressão e identidade, mas também estabelece um solo fértil para a discriminação sistêmica, no qual certos grupos são desfavorecidos em função de que lhes são características inerentes ou atribuídas. O reconhecimento facial, os sistemas de pontuação de crédito e os processos de seleção automatizados emergem como áreas-chave em que o racismo algorítmico se faz presente, revelando as consequências práticas de vieses incorporados em sistemas tecnológicos.

5.1. Viés Algorítmico e Racismo Algorítmico

As expressões “racismo algorítmico” e “viés algorítmico” são formas atuais de representar a prática do racismo atrelado não apenas a aspectos éticos relacionados ao uso das tecnologias de informação e muito menos como algo independente, criado pelo desenvolvimento da inteligência das máquinas, mas como constituinte de um fenômeno contemporâneo de práticas de violência racial. Não se trata, porém, de meras expressões, pois possuem carga conceitual. Assim sendo, o conceito de viés algorítmico refere-se à tendência que os sistemas baseados em algoritmos possuem de reproduzir ou amplificar preconceitos existentes nas fontes de dados a partir das quais são treinados, podendo manifestar-se de diversas formas, em decisões automatizadas por algoritmos de inteligência artificial, seja de maneira autônoma, seja com assistência humana.

Sua utilização em diversas esferas da sociedade contemporânea, como justiça criminal, mercado de trabalho, finanças, saúde e interação social, é uma questão sensível, cuja ocorrência resulta em viés prejudicial a grupos historicamente marginalizados, reconfigurando e perpetuando estereótipos sociais.

[13] Refere-se à responsabilidade e à obrigação de responder por ações e decisões, especialmente em contextos em que há impacto significativo sobre outros. Em discussões sobre racismo algorítmico, o conceito de *accountability* ganha uma importância especial, pois envolve a responsabilidade de garantir que algoritmos não perpetuem ou amplifiquem desigualdades raciais (BAROCAS & SELBST 2023).

Apesar dessa visão geral, há a divisão de viés algorítmico em dois tipos, implícito e explícito (AMODIO, 2014), cuja distinção é relevante para compreender as formas variadas pelas quais os sistemas algorítmicos podem perpetuar desigualdades. O explícito refere-se a preconceitos diretamente embutidos nos modelos por meio de decisões conscientes e intencionais durante o desenvolvimento. Esse tipo pode surgir, por exemplo, quando um programador ajusta manualmente parâmetros de um modelo para favorecer ou desfavorecer determinados grupos, sendo esse um problema mais fácil de identificar e corrigir, pois há uma ação deliberada envolvida.

Já o implícito emerge de forma não intencional, frequentemente derivando de padrões históricos ou sociais presentes nos dados de treinamento. A própria composição dos dados gerados por relações humanas e processos históricos reflete e perpetua as hierarquias sociais vigentes. Por exemplo, se um algoritmo de recomendação de emprego for treinado com dados históricos, em que há sub-representação de certos grupos étnicos ou de gênero em determinadas profissões, é provável que as recomendações futuras também reflitam essa sub-representação.

O racismo algorítmico, por sua vez, é uma manifestação específica desse tipo de viés, em que os preconceitos raciais são codificados em sistemas tecnológicos, por meio da seleção de conjuntos de dados que não representam adequadamente todos os grupos raciais ou que historicamente refletem desigualdades raciais. Como resultado, os algoritmos podem exacerbar a marginalização de grupos já vulneráveis, perpetuando um ciclo de exclusão e desigualdade.

As fontes nos dados de treinamento refletem frequentemente as desigualdades presentes na sociedade, e esses vieses podem surgir em várias etapas do processo de coleta e preparação dos dados, influenciando os resultados dos algoritmos de *machine learning*, que, conforme Domingos (2017), são projetados para identificar padrões e fazer previsões com base nos dados fornecidos. Se estes contêm vieses implícitos, os algoritmos de *machine learning* aprendem tais vieses e os aplicam nas recomendações futuras. Uma recomendação de músicas, por exemplo, pode favorecer gêneros predominantemente consumidos por um grupo demográfico específico, marginalizando outros grupos.

Outra fonte significativa é o *feedback loop*, em que as recomendações influenciam o comportamento do usuário e, subsequentemente, os dados coletados para refinar o algoritmo. Se um sistema recomendar de modo reiterado certos tipos

de conteúdo, os usuários podem consumir mais desse conteúdo, reforçando os padrões existentes e exacerbando o viés original. Segundo Cozman e Kaufman (2022, p.201), se os dados utilizados para treinar um algoritmo são desequilibrados, se os dados disponíveis para treinar modelos algorítmicos não são representativos da população total ou refletem desigualdades sociais, o modelo aprenderá e replicará essas disparidades.

Para Silva (2022a, pp.26-27), o racismo algorítmico é a reprodução e a intensificação de preconceitos raciais por meio de algoritmos e tecnologias digitais e se manifesta na forma como sistemas automatizados classificam e tratam informações, favorecendo grupos privilegiados e marginalizando minorias. Além de discursos explícitos de ódio, o racismo algorítmico inclui práticas sutis, como microagressões, que estão integradas às estruturas de dados e interfaces tecnológicas. Dessa forma, ele reflete e reforça as desigualdades sociais e as hierarquias raciais existentes na sociedade. Silva (2020, pp.137-138) também sistematiza diversas ocorrências passíveis de interpretação como microagressões digitais, fundamentando-se em um trabalho de mapeamento ainda em desenvolvimento.

A Tabela 1, sistematizada a partir de Silva (2022a), resume 12 casos de microagressões digitais identificados em sistemas como plataformas de anúncios, mecanismos de busca de imagens, ferramentas de processamento de linguagem, visão computacional e *chatbots*:

Tabela 1: Lista de Casos de Racismo Algorítmico mapeados

Caso de Racismo Algorítmico	Microagressões	Categoria
*Sistema do Google permite empresas exibirem anúncios sobre crime especificamente a afro-americano**s*** (SWEENEY, 2013).	Suposição de criminalidade	Microinsultos
Resultados no Google Imagens apresentam hiperssexualização para buscas como "garotas negras" (NOBLE, 2013, 2018).	Exotização; negação de Cidadania	Microinsultos
Facebook esconde manifestações contra violência policial racista (TUFEKCI, 2014).	Negação de realidades raciais	Microinvalidações
Google Photos marca fotos de jovens negros com a tag "Gorila".	Negação de Cidadania	Microinsultos
Chatbot da Microsoft torna-se racista em menos de um dia.	Diversas	Microinsultos
Robôs conversacionais de startups não encontram face de mulher negra; sistemas de visão computacional erram gênero e idade de mulheres negras (BUOLAMWINI, 2018).	Negação de cidadania; exclusão e isolamento	Microinvalidações
Mecanismos de busca de bancos de imagens invisibilizam famílias e pessoas negras.	Negação de realidades raciais	Microinvalidações; desinformação

App que transforma selfies equipara beleza à brancura.	Exotização; exclusão e isolamento	Microinsultos; microinvalidações
APIs de visão computacional confundem cabelo negro com perucas (MINTZ *et al.*, 2019).	Exotização	Microinsultos; microinvalidações
Ferramentas de processamento de linguagem natural possuem vieses contra linguagem e temas negros.	Patologização de valores culturais	Deseducação
Análise facial de emoções associa categorias negativas a atletas negros.	Suposição de Criminalidade	Microinsultos
Twitter [atual *X*] *decide não banir discurso de ódio nazista/supremacista branco para não afetar políticos republicanos*.	Negação de realidades raciais; Exclusão	Deseducação; desinformação

Fonte: Inspirada em Silva (2022a).

Esses exemplos ilustram como o racismo algorítmico pode ser compreendido através das microagressões, cujos efeitos são amplificados ou modificados pelas particularidades dos ambientes digitais.

Portanto, a incorporação de vieses raciais em sistemas algorítmicos não apenas perpetua desigualdades existentes, mas também institucionaliza novas formas de discriminação que são mascaradas pelas aparentes objetividade e neutralidade tecnológica. Porém, a opacidade dos modelos algorítmicos e a falta de transparência nas decisões automatizadas dificultam o diagnóstico e a correção de vieses. Para O'Neil (2020, p.46), a falta de transparência nos modelos torna o modo de operação inacessível para aqueles que não são especialistas em computação.

Assim, mesmo em casos de falhas, as decisões dos algoritmos não podem ser questionadas, contribuindo para o agravamento das desigualdades. As estruturas subjacentes não apenas facilitam a emergência e a manutenção do viés, mas também moldam as suas consequências. Se esses sistemas são treinados com dados históricos que refletem práticas discriminatórias, eles, evidentemente, continuam a marginalizar comunidades que já estão em desvantagem, ao invés de proporcionar uma plataforma para equidade.

Silva (2022a, pp.37-38) assevera que as categorias usadas nas plataformas digitais e as interfaces de usuário desempenham um papel crítico na maneira como as informações são apresentadas e apreendidas. Tais categorias não são, de modo algum, neutras, resultando em distorções a partir das quais estigmas podem ser multiplicados. As interfaces do usuário, em última instância, desenhadas para auxiliar na interação, impactam no consumo das informações. Questões de design como tamanho de letra ou *layout*, por exemplo, podem causar a exposição positiva a

conteúdos que estão carregados de estereótipos que os usuários internalizam como representações normativas.

Além disso, algoritmos de recomendação, que operam baseados em dados de comportamento, podem perpetuar a invisibilidade, marginalizando conteúdos que não representam efetivamente certos grupos, em um ciclo vicioso. Segundo Silva (2022a, pp.47-48), o fenômeno dessas “bolhas de filtro” resulta na exposição limitada a informações diversas, reforçando preconceitos e dificultando a inclusão de vozes marginalizadas. Esse fenômeno significaria conclusões uniformizadas extraídas das informações acessadas. Como resultado, os serviços expõem o usuário principalmente a determinados tipos de conteúdo com os quais concorda, diminuindo significativamente a diversidade de opiniões e experiências, que é um dos indicadores da diversidade.

As práticas de moderação e algoritmos por trás das plataformas digitais influenciam na forma como o conteúdo é apresentado aos usuários. Não decidem apenas quais informações são priorizadas e as mais visíveis para um público maior, mas também decidem sobre a experiência do usuário, consistindo em quais tipos de informações são priorizados para visualização ou são ocultados. Essa moderação lida com a identificação e o controle de informações consideradas prejudiciais, mas a ordenação e a recomendação personalizadas vão além, afetando diretamente a diversidade de perspectivas disponíveis.

6 RACISMO ALGORÍTMICO COMO UMA NOVA ETAPA DO RACISMO COMO FORMAÇÃO DISCURSIVA

Este capítulo objetiva apresentar, sob o ponto de vista teórico de Michel Foucault, relativamente ao conceito de formação discursiva, o racismo algorítmico como uma nova representação desse conceito, pela compreensão de que o racismo, como prática social e discursiva, ao longo do tempo, sofreu transformações significativas dentro de novos contextos tecnológicos, nos quais os sistemas computacionais se tornaram agentes na produção, na reprodução e na disseminação de desigualdades raciais.

A partir da visão foucaultiana de que as formações discursivas legitimam discursos como instrumentos que estabelecem “regimes de verdade” e que por meio desses discursos o biopoder atua, criando medidas regulatórias que governam os corpos e a vida das populações em suas diversas dimensões, compreende-se o racismo algorítmico como uma nova formação discursiva, organizada em torno de normas e regras específicas, determinada por condições de existência, manutenção e modificação de enunciados em um campo discursivo próprio.

Foucault (2019, p.8) observa que, tradicionalmente, a história se dedicava a preservar os monumentos[14] do passado, convertendo-os em documentos e interpretando os vestígios que, muitas vezes, não eram verbais ou expressavam algo diferente do que aparentavam. Atualmente, a história é responsável por transformar documentos em monumentos, revelando uma série de elementos que precisam ser isolados, agrupados, relacionados e organizados em conjuntos, em que antes se buscava decifrar os rastros deixados pelos homens e compreender profundamente o que eles foram. Para o autor, monumento não é apenas uma estrutura física ou um objeto de valor histórico, mas uma forma de dar *status* e significado a um conjunto de documentos e vestígios do passado.

O conceito de monumento, portanto, implica uma reflexão sobre como a história é construída e como os discursos são formados. Foucault sugere que, ao invés de

[14] Refere-se à transformação do documento histórico em um objeto de análise que não apenas registra o passado, mas também revela as condições de possibilidades dos discursos. Os monumentos são tratados como vestígios que permitem a reconstrução de discursos e práticas de uma época. Eles não são apenas registros passivos, mas sim ativos na produção de sentidos e na manutenção de relações de poder, controle e memória. Essa abordagem rompe com a ideia tradicional de história como continuidade e enfatiza a descontinuidade, destacando os documentos como vestígios de formações discursivas específicas.

perseguir uma narrativa linear e coesa, é essencial examinar as práticas discursivas que originam esses monumentos. Devem-se levar em conta as relações de poder e as condições sociais que moldam a produção do conhecimento, fazendo-se necessária uma análise mais profunda das práticas discursivas e das relações sociais que moldam a memória coletiva e a construção do conhecimento histórico.

Dentro dessa perspectiva, interessa para a presente pesquisa identificar as circunstâncias que levam ao surgimento de enunciados enviesados, as normas que os regem e as dinâmicas de poder que moldam sua criação. Sob essa ótica, o método arqueológico desvia-se da narrativa histórica convencional, focando nas descontinuidades e nos padrões que possibilitam o aparecimento e a evolução dos enunciados. Esse aspecto é especialmente relevante para analisar as condições de possibilidade por trás da produção de formas específicas de saber e fornece, assim, uma compreensão mais detalhada dos processos de inclusão e exclusão subjacentes às práticas discursivas.

Ao se aplicar esse conceito, podem-se mapear as redes de poder que entram em diálogo com o saber e entender, portanto, como as relações de poder são articuladas através das diferentes linhas de força. Em um contexto de transformação acelerada provocada pela disseminação do uso dos ambientes virtuais, a abordagem arqueológica foucaultiana fornece uma análise que possibilita investigar como essas novas tecnologias interagem com discursos preexistentes, ao mesmo tempo em que são apresentadas para a constituição de novas epistemes[15].

O conceito de biopolítica de Foucault (1998), igualmente fundamental nesta discussão, demonstra como certas formações discursivas, como o saber médico, o religioso ou o pedagógico, por exemplo, estão associadas a práticas institucionais que regulam não apenas o conhecimento, mas também o comportamento e a subjetividade. Esses elementos fundamentais – regularidade dos enunciados, condições de possibilidade e interseção com o poder – são cruciais para entender como as formações discursivas operam em diferentes campos. O discurso médico, por exemplo, se estrutura com base na organização dos saberes sobre o corpo, a

[15] Por *episteme*, entende-se, na verdade, o conjunto das relações que podem unir, em uma dada época: as práticas discursivas que dão lugar a figuras epistemológicas, a ciências, eventualmente a sistemas formalizados; o modo segundo o qual, em cada uma dessas formações discursivas, se situam e se realizam as passagens à epistemologização, à cientificidade, à formalização; a repetição desses limiares que podem coincidir, ser subordinados uns aos outros ou estarem defasados no tempo; e as relações laterais que podem existir entre figuras epistemológicas ou ciências, na medida em que se prendam a práticas discursivas vizinhas mais distintas (FOUCAULT, 2019. p.230-231).

doença e a saúde, a partir de certas práticas e instituições (FOUCAULT, 2021). De forma similar, outras formações discursivas, como a da economia ou a da religião, estabelecem suas próprias verdades e regras de funcionamento, sempre em diálogo com o contexto histórico e social em que se inserem (FOUCAULT, 2008).

O racismo algorítmico pode ser entendido como uma formação discursiva contemporânea em que os algoritmos, como práticas tecnológicas, refletem e perpetuam as relações de poder existentes, incluindo as estruturas racistas. Assim, o racismo é codificado e disseminado através dessas novas práticas discursivas digitais, perpetuando a discriminação em um novo domínio – o digital.

Na obra *A Arqueologia do Saber* (1969/2019), Foucault explora o conceito de formação discursiva, questionando a possibilidade de atribuir uma unidade ao discurso. Ele afirma que não é possível estabelecer essa unidade com base nos objetos do discurso, como o discurso sobre a loucura, pois eles variam e são tratados diferentemente ao longo do tempo. Para o autor, a unidade pode ser encontrada nas regras que permitem a emergência de diferentes objetos dentro de um discurso, não reside, portanto, na forma dos enunciados. Ao investigar a formação dos conceitos, ele observa que, mesmo dentro de um único discurso, como o das ciências humanas, os conceitos podem ser conflitantes e criar um campo de disputa. Por fim, ao considerar os temas dos discursos, Foucault percebe que é mais produtivo observar a dispersão das escolhas temáticas e as regularidades que governam os enunciados dentro de uma formação discursiva do que buscar uma unidade linear.

Portanto, cada formação discursiva é produto de condições históricas específicas. Em um contexto contemporâneo, os discursos tecnológicos moldam a percepção e a utilização de novas tecnologias. Por exemplo, a maneira como se fala sobre inteligência artificial, big data e algoritmos é influenciada por formações discursivas que incluem conceitos de eficiência, progresso e inovação, bem como preocupações com privacidade, segurança e ética. O racismo algorítmico não é apenas um reflexo de preconceitos preexistentes; ele constitui uma nova maneira de estruturar as relações de poder na sociedade. Observa-se essa relação na forma como o saber, enquanto conhecimento técnico, se traduz em sistemas que perpetuam o poder por meio, por exemplo, das desigualdades raciais. O racismo algorítmico como formação discursiva organiza as estruturas de poder subjacentes à criação e à

implementação de tecnologias aparentemente neutras. Os algoritmos, como nova forma de poder-saber, moldam nossa compreensão da realidade, uma vez que não apenas processam dados, mas também criam “verdades” e influenciam comportamentos, refletindo e perpetuando preconceitos existentes.

Os sistemas algorítmicos disciplinam as subjetividades decidindo quem merece atenção, recursos e reconhecimento social e quem deve ser mantido à margem. Logo, seu uso enviesado afeta não apenas as estruturas de poder e conhecimento, mas também a formação de subjetividades, práticas de governança e a própria constituição do que entendemos como “raça” na era digital. Essa formação discursiva é constituída por:

a) Objetos – algoritmos, bases de dados, sistemas de inteligência artificial que reproduzem ou amplificam vieses raciais;

b) Modalidades enunciativas – relatórios técnicos, estudos acadêmicos, debates públicos e políticas corporativas que abordam o tema;

c) Conceitos – viés algorítmico, discriminação automatizada, equidade tecnológica etc.; e

d) Estratégias temáticas: discussões sobre ética na IA, regulamentação tecnológica, justiça social digital, entre outros.

Diante de um cenário de popularização da cultura digital, é flagrante o impacto da internet na formação e na percepção de grupos identitários específicos, sobretudo em questões relacionadas à raça e à etnia. A ideia de uma identidade fluida, modelada pelas interações on-line, emergiu num contexto de opções limitadas de comunicação digital, uma representatividade restrita a pesquisadores, e a uma crença, reforçada por um discurso dominante, acerca da neutralidade da rede.

No entanto, pesquisadores e ativistas destacam que as tecnologias digitais e a ideologia dominante no setor tecnológico são moldadas por conceitos de supremacia racial. Além disso, estratégias de diminuição dos debates sobre racismo em mídias sociais são observadas, como a descontextualização da relevância acerca de aspectos da cultura negra. Essas dinâmicas indicam que o racismo algorítmico se apresenta como uma nova fase na evolução discursiva do racismo, influenciando diretamente a formação e na percepção de identidades raciais e étnicas na era digital.

O surgimento dessa nova formação discursiva se valeu bastante da noção de plataformização[16] da internet, que descreve a expansão de ambientes digitais, como o Facebook, que se transformaram em *hubs*[17] centrais de dados e valor, dominados por algumas corporações. Essa transformação está alinhada com a concepção de Helmond (2019) sobre plataformas de mídia social como infraestruturas tecnológicas sobre as quais outros podem construir, visando expandir sua presença na *web*. Segundo Srnicek (2017), plataformas são infraestruturas digitais que possibilitam a interação entre dois ou mais grupos. Elas se posicionam como intermediárias que aproximam consumidores, prestadores de serviço, vendedores, fornecedores, empresas de propaganda, servindo como uma base para que suas atividades ocorram e, assim, possam se privilegiar do acesso a esses dados.

À medida que a dataficação[18] e a mediação das atividades humanas se intensificam, observa-se uma concentração de dados e capital que contrasta com os discursos promovidos pelas plataformas digitais sobre liberdade e igualdade nas relações sociais e digitais. Essa contradição é particularmente evidente no contexto das interações on-line e no acesso a serviços digitais, em que se promete democratização e inclusão, mas, na prática isso não efetivando, daí os algoritmos, como sequências finitas de instruções projetadas para executar tarefas em sistemas computacionais, aprimorados com processos de inteligência artificial, remodelarem, intensificarem e atualizarem formas de segregação racial.

Essa prática nos aponta para uma nova fase na formação discursiva do racismo, em que relações entre usuários e plataformas digitais, o acesso a oportunidades econômicas e sociais mediadas por tecnologia, a representação e o tratamento de grupos raciais e étnicos dentro desses sistemas informacionais se dão

---

[16] Helmond (2015) define as plataformas como infraestruturas que moldam a maneira como os sujeitos se comunicam, consomem e interagem com o mundo digital. Desdobramento do conceito de plataforma, a plataformização, segundo a autora, é um processo de emergência e consolidação das plataformas como modelo econômico e infraestrutural dominante das redes sociais on-line (HELMOND, 2015).

[17] O Facebook é considerado um *hub* em virtude de seu tamanho, alcance, conectividade, influência, centralização e inovação. A plataforma se apresenta como ponto de encontro para pessoas ao redor do mundo e um poderoso canal de comunicação, compartilhamento de informações e influência social. Vale salientar que essa *big tech* recebe críticas por questões relacionadas à privacidade, à manipulação de dados e à desinformação.

[18] Segundo Mayer-Schönberger & Cukier (2013) e Zuboff (2021), a dataficação é o processo de transformar aspectos da vida social em dados quantificados, revolucionando a maneira como as pessoas vivem, trabalham e pensam. Consiste na coleta e na análise de grandes volumes de dados sobre o comportamento humano. Esse processo é utilizado por empresas de tecnologia para criar um sistema de “capitalismo de vigilância”, a partir da exploração de dados dos usuários para fins lucrativos, ameaçando a democracia e a liberdade individual.

de forma enviesada. Trata-se de uma prática de racialização e exclusão social, em que certas populações são preteridas e, consequentemente, sujeitas a formas de controle e vigilância mais intensas.

Para uma melhor compreensão dessa conjuntura, destaca-se a relevância do conceito de biopoder, já abordado anteriormente. Como uma dimensão do biopoder, para Foucault (2021b, 1998), a biopolítica é outro conceito importante, pois revela o poder sobre a vida, enfatizando como as práticas governamentais se infiltram na biologia dos corpos humanos para regulamentar as populações. Esse enfoque na gestão da vida e na soberania sobre a morte introduz uma perspectiva na qual o racismo e a exclusão social são examinados como consequências diretas das estratégias biopolíticas.

A institucionalização do racismo sob a tutela da biopolítica é evidente em várias políticas e práticas, incluindo leis de imigração, políticas de saúde pública e sistemas de justiça criminal. Essas instâncias demonstram como o racismo, mediado por práticas biopolíticas, se infiltra nas estruturas institucionais, perpetuando desigualdades e hierarquias raciais. Foucault (2019) argumenta que toda formação discursiva está associada a práticas institucionais que operam para sustentar as relações de poder dentro de uma sociedade. No caso do racismo algorítmico, essas instituições são, em grande parte, empresas de tecnologia, plataformas digitais e governos que adotam e utilizam esses sistemas para controlar e administrar populações.

No contexto da sociedade contemporânea, a partir da análise das relações entre poder e tecnologia, é possível observar uma atualização do biopoder, em que tecnologias digitais e algoritmos já desempenham um papel ativo na gestão da vida, pela capacidade de coletar, analisar e utilizar grandes volumes de dados pessoais para influenciar comportamentos e decisões individuais.

Esse conjunto de argumentos permite afirmar que o racismo algorítmico se constitui como uma formação discursiva. Entretanto, para o melhor delineamento dessa afirmativa, há que se buscar na definição de racismo epistêmico um *insight* sobre o racismo como formação discursiva, com o propósito de se estabelecer paralelos com as teorias sobre o biopoder, que dá sustentação ao objeto desta pesquisa. O racismo epistêmico refere-se à forma como o racismo está enraizado nos processos de produção e legitimação do conhecimento. Para Grosfoguel (2006, p.41), trata-se de um dos racismos mais invisibilizados no sistema-mundo

capitalista/patriarcal/moderno/colonial. O racismo em nível social, político e econômico é muito mais reconhecido e visível que o racismo epistêmico, que privilegia as políticas identitárias dos brancos ocidentais. Grosfoguel (2006, p.44) argumenta que os diversos fundamentalismos, mesmo os eurocêntricos, baseiam-se num pressuposto comum: a defesa de um monopólio cognitivo como único caminho válido para a verdade e a universalidade.

O racismo epistêmico consiste, portanto, na desvalorização e na invisibilidade de saberes e experiências dos negros, epistemologia da desumanização utilizada como ferramenta para justificar a dominação e a exploração que marcam a história do povo negro. Essa violência reforça a desvalorização dos saberes, perpetuando um ciclo de opressão e marginalização, evidenciado por desigualdades sociais, violência física, simbólica, epistemológica e pela desumanização. Mbembe (2018b) argumenta que esse poder se manifesta de forma específica sobre os corpos negros, através da escravidão, do colonialismo, do *apartheid* e da violência policial.

Santos (2010) assevera que esse tipo de racismo pode ser mais sutil do que as justificações biológicas ou culturais, mas suas consequências são igualmente prejudiciais, perpetuando relações de poder e opressão:

> O racismo – sob a forma de esquecimento da condenação, racismo epistêmico e muitas outras formas – está mais disseminado do que frequentemente se pensa. [...]. Para além das justificações biológicas de racismo, ou das justificações baseadas em diferenças de cultura ou maneiras de estar, é possível encontrar em algumas tendências influentes do pensamento ocidental uma justificação ontológica e epistemológica mais sutil. As consequências são nefastas, uma vez que a fusão de espaço e raça está por trás de concepções militares e imperiais da espacialidade, que tendem a dar um novo significado à formulação clássica de Santo Agostinho acerca das cidades terrenas e divinas[19] (SANTOS, 2010, p.367).

O racismo epistêmico é uma realidade presente em diversas esferas da sociedade, respaldando práticas que perpetuam o racismo institucional e o genocídio antinegro como necropolítica de Estado. Mbembe (2018b, p.146) assevera que as formas contemporâneas que subjugam a vida ao poder da morte (necropolítica) reconfiguram profundamente as relações entre resistência, sacrifício e terror. O autor afirma que há uma interconexão entre o racismo epistêmico e o biopoder, revelando um arranjo cruel que perpetua a marginalização e a violência contra a população negra.

---

[19] A diferença entre a Cidade de Deus e a Cidade Terrena dos Homens traduz-se na divisão entre as cidades imperiais dos deuses humanos e as cidades dos condenados. Infelizmente, a busca de raízes na Europa e as geopolíticas racistas costumam andar de mãos dadas (SANTOS, 2010, p.367).

Foucault, em suas análises acerca do poder e da sociedade, descreve uma transição significativa para o que ele denomina sociedade disciplinar. Essa mudança marca uma era na qual o poder não se manifesta meramente por intermédio da força ou da soberania, mas se infiltra na malha social de maneira mais sutil e abrangente. A sociedade disciplinar caracteriza-se pelo surgimento de instituições e práticas que visam à ordenação e ao controle dos sujeitos por meio de uma série de técnicas disciplinares (FOUCAULT, 2020, pp.289-291). Essas técnicas, segundo Foucault, operam no nível mais íntimo e pessoal, atuando diretamente sobre os corpos com o objetivo de moldar, treinar e corrigir comportamentos.

Nesse contexto, o poder se torna mais capilarizado, espalhando-se por toda a estrutura social e permeando as relações cotidianas. As disciplinas, como Foucault as denomina, emergem em variados contextos, desde escolas e exércitos até hospitais e prisões, instituindo um regime de vigilância constante e avaliação normativa. Essa nova forma de poder busca não apenas punir, mas principalmente prevenir desvios, antecipar infrações e integrar o corpo social de maneira produtiva e útil.

A tecnologia não apenas atua como mediadora de decisões, mas também como uma extensão do poder disciplinar e biopolítico que regula vidas, identidades e subjetividades. Essa articulação entre tecnologia e poder reflete o que Foucault (2021) chamou de “governamentalidade”, ou a arte de governar populações por meio de regimes de verdade, que no contexto contemporâneo estão cada vez mais mediatizados pelas lógicas algorítmicas e digitais, nas quais a regulação de algoritmos se torna um novo campo de intervenção política, criando formas de governamentalidade.

Trata-se de um conjunto de práticas, técnicas e racionalidades que objetivam gerir e controlar a vida da população, não se restringindo apenas ao governo político formal, mas abrangendo uma série de dispositivos e mecanismos que atuam na regulação dos corpos, nas suas condutas e nos seus comportamentos. Esses dispositivos incluem não apenas leis e instituições, mas também práticas disciplinares, técnicas de segurança e estratégias de vigilância.

Segundo Foucault (2008), o poder se insere nas práticas cotidianas, moldando subjetividades e estabelecendo normas de conduta, não apenas de forma coercitiva, mas também por meio de processos de normalização e regulação. Para Foucault, governamentalidade é:

> o conjunto constituído pelas instituições, os procedimentos, análises e reflexões, os cálculos e as táticas que permitem exercer essa forma bem específica, embora muito complexa, de poder que tem por alvo principal a população, por principal forma de saber a economia política e por instrumento técnico essencial os dispositivos de segurança. Em segundo lugar, por "governamentalidade" entendo a tendência, a linha de força que, em todo o Ocidente, não parou de conduzir, e desde há muito, para a preeminência desse tipo de poder que podemos chamar de "governo" sobre todos os outros – soberania, disciplina – e que trouxe, por um lado, o desenvolvimento de toda uma série de aparelhos específicos de governo e, por outro lado, o desenvolvimento de toda uma série de saberes. Enfim, por "governamentalidade", creio que se deveria entender o processo, ou antes, o resultado do processo pelo qual o Estado de justiça da Idade Média, que nos séculos XV e XVI se tornou o Estado administrativo, viu-se pouco a pouco "governamentalizado" (2008, pp.143-144).

A governamentalidade representa um tipo de poder que se preocupa primordialmente com a gestão da população, abrangendo uma vasta gama de práticas e raciocínios que visam o bem-estar do corpo social, a eficiência econômica e a segurança do Estado. Essa forma de governar é calculista e reguladora, orientada não apenas pelo desejo de dominar, mas pela necessidade de otimizar e gerir a vida dos cidadãos de maneira eficaz.

Na era digital, a governamentalidade não se refere apenas à maneira como o Estado exerce controle sobre os cidadãos, mas também ao modo como corporações, instituições e outros atores utilizam as tecnologias de informação para exercer uma forma de poder que é, ao mesmo tempo, dispersa e concentrada. A capacidade de influenciar a opinião pública, monitorar comportamentos e até mesmo prever ações futuras através da análise de dados coloca em questão as fronteiras tradicionais entre o público e o privado, o Estado e o sujeito.

Essa capacidade de influenciar e determinar a ação individual, baseada na coleta massiva de dados, alinha-se com a ideia foucaultiana de governamentalidade, na qual o poder se exerce de maneira difusa, através da gestão da vida individual e da população como um todo. É a partir dessa materialidade tecnológica, oriunda da mineração de dados pessoais, que são reproduzidas dinâmicas de opressão já cristalizadas em outros ambientes sociais, por meio da geração e do gerenciamento de informações que não são disponibilizadas de maneira democrática. Ao contrário, são inúmeros dados sobre o corpo social sob a tutela de poucas pessoas com interesses, inclusive, de exclusiva lucratividade.

Um dos instrumentos de governamentalidade que se pôde identificar ao longo desta pesquisa é justamente o racismo algorítmico, que, segundo Noble (2021) e Benjamin (2019), cria regularidades ao estruturar as decisões algorítmicas em

sistemas de reconhecimento facial, recomendações em mecanismos de busca, ou até no monitoramento de segurança pública, em que os algoritmos decidem quem é suspeito, quem recebe recursos e quais vidas são valorizadas ou descartadas, seguindo uma lógica de exclusão e controle, criando condições de possibilidade que moldam as práticas sociais e culturais em torno do que é considerado verdade ou visível no mundo digital.

Ao invisibilizar e marginalizar certos grupos, esses sistemas estabelecem uma ordem de verdade que determina quem é “legítimo” e quem é passível de vigilância, exclusão ou discriminação de modo automatizado. Nessa formação discursiva, os algoritmos operam sob lógicas de mercado que priorizam determinados tipos de conteúdo, frequentemente marginalizando ou distorcendo a representação de grupos raciais minoritários. Eubanks (2018) assevera que a automação e os algoritmos são utilizados para monitorar e controlar populações marginalizadas, especialmente as minorias raciais. Nesse sentido, assegura a autora:

> Os partidários das abordagens automatizadas e algoritimizadas ao serviço público frequentemente descrevem a nova geração de ferramentas digitais como “inovadoras”. Eles nos dizem que as big data chacoalham as burocracias inflexíveis, estimulam soluções inovadoras e aumentam a transparência. Mas quando focamos nos programas especificamente destinados aos pobres e às pessoas da classe trabalhadora [...], Ele é simplesmente uma expansão e o prosseguimento das estratégias punitivas e moralistas de gerenciamento da pobreza que nos acompanham desde os anos 1820 (EUBANKS, 2018, p.37).

Nesse contexto, outra categoria importante para o entendimento de que o racismo algorítmico se constitui como uma formação discursiva é a compreensão do racismo algorítmico como discurso, que tem seu campo próprio de poder e saber, estabelecendo fronteiras entre quem pode se beneficiar da tecnologia e quem é prejudicado por ela, criando formas de exclusão que se perpetuam por meio de decisões automáticas, reforçadas por sistemas opacos e complexos, os quais invisibilizam grupos racializados e favorecem outros de maneira sistemática.

Na perspectiva foucaultiano, é possível conceber essa prática como discurso por ela organizar e regular saberes, definindo quem tem visibilidade e acesso a certos recursos e quem é marginalizado por essas tecnologias. Porém, o discurso não é apenas uma representação da realidade, mas uma prática que constitui a própria realidade, moldando a forma como os sujeitos percebem o mundo e interagem com ele, em compasso com o que afirma Orlandi (2009).

Para Foucault (2021), uma formação discursiva se estabelece quando há uma regularidade em um campo de saber que não só estrutura os discursos, mas também delimita os sujeitos, as práticas e as instituições que têm autoridade para definir o que é verdadeiro. O racismo algorítmico atua como uma formação discursiva ao definir quem é visível e reconhecido pelas tecnologias e quem permanece marginalizado por elas, visto que as formações discursivas contemporâneas se expandem e transformam, adaptando-se aos novos campos de saber, como o digital, no qual o controle biopolítico ocorre de forma mais sofisticada e imperceptível. Para Foucault (2019), as formações discursivas não apenas definem o que pode ser dito, mas também o que é considerado verdadeiro ou falso em um determinado campo. Sob o pretexto da “neutralidade”, os algoritmos, de igual modo, moldam o que pode ser dito, pensado e, mais importante, decidido por sistemas automatizados, ao processarem e disseminarem dados enviesados. Eles se tornam instrumentos de poder que não apenas reproduzem desigualdades raciais, mas também criam outras formas de controle social.

O racismo algorítmico como discurso não se limita a ser uma simples aplicação tecnológica. Ele opera como uma formação discursiva que gera “regimes de verdade” nas relações sociais contemporâneas. Essa produção de verdade é perpetuada pela aceitação generalizada de que os algoritmos são objetivos e imparciais, quando na realidade eles reproduzem desigualdades sociais, atuando como mecanismos de poder e controle sobre corpos racializados.

Dessa forma, o discurso racista algorítmico não apenas atualiza as práticas racistas do passado, mas também as reconfigura para o contexto digital, ao produzir e disseminar verdades que consolidam novas formas de discriminação. Assim a formação discursiva da medicina regulou a saúde dos corpos e os padrões de beleza, e a formação discursiva econômica moldou as relações de produção e trabalho. Portanto, reiterando, o racismo algorítmico define como determinados corpos e identidades são excluídos ou marginalizados no ambiente digital.

Na perspectiva do racismo algorítmico, pode-se ver como os discursos produzidos por algoritmos reconfiguram a visão da sociedade sobre raça e identidade. Eles são internalizados pelos sujeitos e se tornam parte da estrutura social, influenciando desde as políticas públicas até as interações cotidianas. Assim, o racismo algorítmico, enquanto uma nova formação discursiva, não apenas reflete as

desigualdades existentes, mas também as amplifica e as solidifica em novas formas de exclusão.

Para Foucault (2021), os discursos têm o poder de definir o que é considerado “normal” dentro de uma sociedade, e essa normatividade é frequentemente construída de maneira a beneficiar certos grupos em detrimento de outros. No caso dos algoritmos, ao serem inseridos em sistemas de decisão automatizados, participam na construção de subjetividades e na definição do que é considerado “normal’ ou “desviante”. No entanto, as decisões automatizadas que parecem ser baseadas em critérios objetivos, de modo sistêmico, escondem vieses incorporados, que são perpetuados de forma sistemática e amplificada, naturalizando a exclusão racial.

Segundo Piovezani (2015), ao apresentarem decisões como objetivas e imparciais, os algoritmos mascaram as ideologias subjacentes. Essa ocultação torna ainda mais difícil a tarefa de contestar o racismo algorítmico, pois ele se apresenta como um fato técnico, e não como uma construção social ideologicamente carregada. A naturalização dessas práticas através do discurso algorítmico cria uma hegemonia discursiva, na qual as desigualdades raciais são invisíveis, mas profundamente enraizadas. Esses sistemas, ao automatizarem a discriminação racial, participam na construção de um novo regime de verdade, em que a exclusão e a marginalização são apresentadas como consequências naturais de processos aparentemente neutros.

7 A DESCRIÇÃO DO *CORPUS* E SUA DISCUSSÃO

Esta pesquisa adota uma abordagem qualitativa, fundamentada na análise do discurso com o objetivo de investigar as manifestações do racismo algorítmico à luz da analítica do biopoder de Michel Foucault. A escolha metodológica justifica-se pela necessidade de compreender como as práticas discursivas, em meios tecnológicos, reconfiguram, reproduzem e atualizam relações de poder e controle sobre corpos racializados, desvelando as estruturas inscritas nesse ambiente digital, em consonância com os conceitos foucaultianos, como formação discursiva, biopolítica, acontecimento e biopoder.

O *corpus* desta pesquisa foi constituído a partir de dois eixos principais: a Linha do Tempo do Racismo Algorítmico e notificações da ferramenta Google Alerts. Quanto à Linha do Tempo do Racismo Algorítmico, uma *Timeline* interativa com um escopo o qual cobre casos e dados de danos e discriminação algorítmica, cuja visualização é atualizada continuamente e que reúne casos e reações relacionados ao racismo algorítmico no período de 2010 a 2024.

Foram selecionados exemplos que ilustram as dinâmicas do racismo algorítmico em diferentes contextos, tais como:

a) Sistemas de branqueamento em filtros de aplicativos e redes sociais, em que se observam os efeitos discursivos e biopolíticos de filtros os quais favorecem e reforçam padrões eurocêntricos;

b) Reconhecimento facial aplicado à segurança pública, em que se investigam as implicações do uso de tecnologias de reconhecimento facial as quais apresentam elevados índices de erro em relação aos negros, o que resulta em criminalização e vigilância desigual;

c) Sistemas de recomendação usados nas redes sociais, em que se exploram os algoritmos que perpetuam estereótipos raciais e segregam usuários com base em critérios racializados;

d) Testes padronizados usados na saúde, em que se examinam os vieses raciais em algoritmos de diagnóstico e tratamento médico que impactam de forma desigual e desproporcional grupos racializados; e

e) Imagens geradas por inteligência artificial, nas quais se analisam as representações raciais produzidas por sistemas de inteligência artificial,

evidenciando como esses dispositivos atualizam práticas de discriminação relativamente a grupos minoritários.

A escolha da *Timeline* sobre racismo algorítmico, elaborada pelo pesquisador Tarcízio Silva, como uma das fontes primárias para a constituição do *corpus* desta pesquisa, justifica-se por algumas questões metodológicas que se adequam aos propósitos desta investigação e a algumas singularidades do objeto de estudo. O autor figura entre os principais pesquisadores brasileiros na área de estudos sobre racismo algorítmico, transparência e responsabilidade em políticas de tecnologia e impactos sociais das práticas algorítmicas. É autor da obra *Racismo Algorítmico: inteligência artificial e discriminação nas redes digitais* (2022) e organizador de livros como *Comunidades, Algoritmos e Ativismos Digitais: olhares afrodiaspóricos* (2020) e *Griots e Tecnologias Digitais* (2022).

A Linha do Tempo sobre Racismo Algorítmico de Tarcízio Silva resulta de um processo criterioso de seleção e organização, é um profícuo trabalho de curadoria e registros de ocorrências sobre racismo, oriundos de pesquisas acadêmicas e estudos publicados em revistas científicas, conferências, reportagens jornalísticas veiculadas de mídias como *The New York Times*, *The Guardian*, *BBC*, *El País*, *Nexo*, *UOL* e *Folha de São Paulo*, relatórios de organizações e ONGs, como AI *Now Institute, Algorithmic Justice League e Human Rights Watch*, publicações técnicas e empresariais, relatórios de empresas de tecnologia e grupos de pesquisa em IA e discussões em redes sociais e blogs, o que confere à fonte um alto nível de credibilidade e rigor acadêmico.

A *Timeline* reúne uma pluralidade de casos, englobando diferentes contextos geográficos, setores da sociedade (como saúde, segurança pública, educação e entretenimento) e tipos de tecnologias (reconhecimento facial, algoritmos de busca, sistemas de recomendação, entre outros). Essa diversidade é essencial para a constituição do *corpus* desta pesquisa, pois apresenta casos selecionados, contextualizados e analisados por especialistas na área, para um estudo que compreende o racismo algorítmico como um fenômeno multifacetado e global, mas com manifestações locais específicas. Essa fonte possibilita, portanto, uma análise comparativa e atualizada dos casos, o que enriquece a pesquisa, à qual foi acrescentada outra ferramenta, corresponde ao segundo eixo referido anteriormente: a busca no Google Alerts.

7.1 Grade da Linha do Tempo do Racismo Algorítmico

A pesquisa *Linha do Tempo do Racismo Algorítmico: casos, dados e reações*, de Silva (2023), faz um registro de ocorrências de discriminação racial por intermédio do uso de algoritmos enviesados nas tecnologias digitais. O mapeamento das ocorrências, apresentado pelo autor, foi agrupado, nesta pesquisa, em 5 eixos temáticos, correspondentes ao período de janeiro de 2010 a dezembro de 2024, para constituição preliminar do *corpus*, a saber: Reconhecimento Facial, Identidade Racial e Visão Computacional; Plataformas Digitais, Mecanismos de Busca e Mídias Sociais; Aplicativos e Serviços; Tecnologias de Vigilância e Ordenação; e Impactos Sociais e Políticos.

7.1.1 Reconhecimento Facial, Identidade Racial e Visão Computacional

No período de março de 2016 a março de 2024, relativamente ao eixo 1 desta pesquisa, foram identificadas 13 ocorrências. São casos, por exemplo, de falhas geradas por algoritmos de reconhecimento facial que não identificam ou distorcem traços de sujeitos negros, incluindo sistemas de vigilância que associam minorias raciais a atividades criminosas de forma desproporcional. Os sistemas de reconhecimento facial da Amazon e da IBM apresentaram erros significativos ao identificar mulheres negras e grupos asiáticos, evidenciando a ineficácia do reconhecimento facial em contextos raciais diversos.

Em fevereiro de 2018, a pesquisa de Joy Buolamwini e Timnit Gebru revelou que aplicativos de reconhecimento facial não conseguiam identificar apropriadamente o gênero e a raça de mulheres negras, apresentando uma precisão muito inferior em comparação com seu desempenho quando se tratava de sujeitos brancos. Os problemas também se estendem a dispositivos vestíveis, como *Fitbit* e *Samsung Gear*, que mostraram menor precisão em peles escuras. Além disso, há registros de algoritmos de saúde que penalizaram pacientes negros com base em dados históricos racistas, ocasionando, por exemplo, sua exclusão da fila de atendimento ou que outros pacientes, de perfil diferente, tivessem atendimento prioritário.

Os dados levantados na *Timeline* de Silva se encontram sistematizados na tabela a seguir:

Tabela 2 – Mapeamento sobre Reconhecimento Facial, Identidade Racial e Visão Computacional

Data	Ocorrência
Março, 2016	*Sistema de exploração de conteúdo enviesa buscas por conhecimento*. O pesquisador mostra como um sistema de exploração de conteúdo pode trazer vieses problemáticos do ponto de vista da biblioteconomia.
Agosto, 2018	*Kairos tira Diversity Recognition do ar*. Startup Kairos reflete sobre impactos dos aplicativos de visão computacional e tira o seu *Diversity Recognition* do ar.
Março, 2019	*Ferramentas de processamento de linguagem natural possuem vieses contra linguagem e temas negros*. A pesquisadora do *MIT Center for Civic Media* mostra problemas em ferramentas de análise e ajustes automatizados de texto.
Janeiro, 2019	*APIs de visão computacional ajustam resultados problemáticos*. Anteriormente auditados, IBM, Microsoft e Megvii melhoram suas taxas de erros. O trabalho mostra, porém, que fornecedores não auditados no estudo original – Amazon e Kairos – apresentam altas taxas de erro.
Março, 2020	*Discriminação algorítmica gera menos choque moral do que discriminação "humana"*. O trabalho sistematiza estudos que mostram assimetria na *percepção*: discriminação considerada algorítmica é menos impactante do que a considerada humana. Além disso, as pessoas tendem a reproduzir de forma mais intensa estereótipos raciais incorporados em sistemas algorítmicos.
Maio, 2021	*Algoritmo exclui estudantes negros e latinos de boas escolas em Nova Iorque*. A reportagem descobriu que algoritmo que distribui estudantes em escolas de ensino médio em Nova Iorque discrimina jovens negros e latinos.
Outubro, 2021	*Startup de chatbots permite discurso racista em personagens*.
Outubro, 2021	*Sistema oferece ética descritiva automatizada, com julgamentos morais racistas e misóginos*. Treinado a partir de pesquisas comportamentais, Mechanical Turk e Reddit, sistema Delphi automatiza violências discursivas.
Maio, 2022	*Canva apresenta somente noivas brancas nas 12 primeiras páginas de resultado*. Designer e pesquisador relatam suas dificuldades em descobrir *conteúdo* com representações de pessoas negras na plataforma.
Outubro, 2022	A startup CharacterAI, que permite a criação de *chatbots* customizados, não *implementou* mecanismos de controle contra práticas discriminatórias.
Dezembro, 2022	*ChatGPT defende tortura contra pessoas do Sudão, Síria, Irã e Coreia do Norte*. O pesquisador solicitou que o *chatbot* escrevesse linhas de código sobre como decidir se alguém pode ser torturado. O código explicitamente abre exceções contra pessoas de quatro países e sugere vigiar mesquitas. < href="https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/">
Fevereiro, 2024	*60% das respostas do OpenAI contêm plágio*. Estudo de ferramenta que identifica plágio analisou respostas do ChatGPT.
Março, 2024	*OpenAI discrimina candidatos a partir de seus nomes*. *OpenAI* usa nomes como *atalho* para discriminar racialmente candidatos de minorias étnico-raciais.

Fonte: Sistematizada a partir de Silva (2022a).

7.1.2 Plataformas Digitais, Mecanismos de Busca e Mídias Sociais

No período de janeiro de 2013 a outubro de 2024, Silva registrou 45 ocorrências de racismo algorítmico relacionados a plataformas e buscadores de resultados algorítmicos. São casos de publicidade discriminatória em plataformas como Google e Facebook, nos quais anúncios de emprego e oportunidades eram direcionados de forma racialmente desigual, estereotipando raça e gênero. São sistemas de anúncios e redes sociais que limitam o alcance de conteúdos de minorias ou direcionam a

publicidade, prática a qual impacta nas oportunidades de trabalho, educação e moradia.

Há, de igual modo, casos de extremismo digital, por intermédio da disseminação de conteúdo extremista no YouTube e seu impacto nas eleições e em movimentos sociais. Em agosto de 2018, a startup Kairos retirou seu aplicativo de reconhecimento de diversidade devido a preocupações com representação racial. Em março de 2019, pesquisas e auditorias revelaram vieses em ferramentas de linguagem e visão computacional, com altos índices de erro entre fornecedores não auditados, como Amazon e Kairos. Além disso, algoritmos do Facebook foram encontrados bloqueando ou censurando postagens de jovens negros sobre racismo, levantando preocupações sobre liberdade de expressão e inclusão digital.

Para uma melhor visualização desses casos contidos na *Timeline* de Silva, atentar para esta tabela:

Tabela 3 – Mapeamento de Plataformas Digitais, Mecanismos de Busca e Mídias Sociais

Data	**Ocorrência**
Janeiro, 2013	*Discriminação em entrega de anúncios*. Pesquisa de Latanya Sweeney identificou que anunciantes no Google conseguem direcionar mensagens a partir de nomes típicos de grupos raciais estadunidenses.
Outubro, 2013	*Busca por "garotas negras" resulta em conteúdo pornográfico*. Busca por "garotas negras" resulta em conteúdo pornográfico. Trabalho de Safiya U. Noble reflete sobre hipervisibilidade de associação do olhar pornográfico sobre garotas negras e latinas como um meio de torná-las ao mesmo tempo "invisíveis" em suas humanidades e complexidades.
Janeiro, 2014	*Hosts brancos cobram mais por locais equivalentes no Airbnb*. O estudo identificou uma disparidade de 12% em valores que anfitriões brancos e não-brancos conseguem na plataforma.
Agosto, 2014	*Facebook esconde manifestações contra violência policial*. Nos *trending topics* do Twitter e de outras mídias sociais, os protestos contra a violência policial racista nos Estados Unidos foram invisibilizados na plataforma.
Julho, 2015	*GooglePhotos taggeou pessoas negras como "gorilas"*. Desenvolvedor Jacky Alciné denuncia que visão computacional do Google Photos marcou pessoas negras como "gorilas", como reportado no *The Guardian*.
Outubro, 2016	*Sistema de anúncios do Facebook permite excluir negros e latinos, prática ilegal*. ProPublica denuncia que sistema de anúncios do Facebook permite excluir por raças (negros, latinos, asiáticos) nos Estados Unidos em categorias como habitação, o que é proibido por lei há décadas. Especialmente curioso é que não permite, entretanto, excluir usuários brancos/caucasianos.
Março, 2017	*Experimento mostra invisibilidade negra nos bancos de imagens*. Experimento da ONG Desabafo Social mostra invisibilidade negra nos bancos de imagens como *Shutterstock*, *GettyImages*, *Depositphotos* e outras.
Junho, 2017	*Regras obscuras do Facebook protegem só homens brancos de discurso de ódio*. Jornalistas descobriram regras do Facebook que explicitamente protegem categorias como "homens brancos", mas não "crianças negras" e outras categorias de discurso de ódio.
Setembro, 2017	*Facebook silencia posts sobre genocídio em Burma*. Regras sobre conteúdo violento atrapalharam visibilidade de denúncias contra "limpeza étnica" em Burma, no Myanmar.

Fevereiro, 2018	*Pilot Parliaments Benchmark Dataset*. Pesquisadoras Joy Buolamwini e Timnit Gebru, do MIT, desenvolveram *dataset* e sistema mais preciso que IBM, Microsoft e Face++.
Abril, 2019	*Algoritmos do Facebook impedem jovens negros de falar sobre racismo na plataforma*. Reportagem mostra relatos de estudantes negros bloqueados na plataforma por escrever contra o racismo.
Abril, 2019	*Estudo mostra que anúncios no Facebook estereotipam raça e gênero*. Anúncio de vagas em empresas de táxi foram entregues predominantemente para usuários negros, enquanto de secretárias foram destinadas para mulheres.
Abril, 2019	*Twitter não vai banir Nazistas algoritmicamente*. Vazou informação que o *Twitter* decidiu não ativar recurso algorítmico para banir nazistas – pois acabaria excluindo políticos Republicanos dos Estados Unidos.
Junho, 2019	*Bancos de imagens hiper ritualizam solidão da mulher negra*. Pesquisadoras da UFRJ e UFRN identificam hiperritualização da representação da solidão da mulher negra em bancos de imagens.
Outubro, 2019	*Buscar "mulher negra dando aula" no Google leva à pornografia*. Google mantém resultados hiperpornográficos em buscas relacionadas a mulheres negras.
Novembro, 2019	*Instagram vê armas e violência em ilustração com garoto negro*. Instagram impede impulsionamento de ilustração por ver "armas" em ilustração inocente de um garoto em cenário de favela.
Abril, 2020	*Google acha que ferramenta em mão negra é uma arma*. Visão computacional da Google confunde instrumento com arma – mas só em mãos negras.
Junho, 2020	*YouTube restringe visibilidade de conteúdos sobre o BlackLivesMatter*. Produtores de conteúdo no YouTube interpelam judicialmente essa plataforma por restringir visibilidade de conteúdo de ativismo político antirracista.
Julho, 2020	*Usuários negros tem 50% mais chance de serem banidos no Instagram*. Estudo descobriu que algoritmo de moderação automática bania usuários negros numa taxa equivalente a 50% a mais do que em casos de pessoas de outras raças.
Agosto, 2020	*Gboard: teclado do Google sugere termos sexuais para a palavra "neguinha"*. Teclado do Google sugere ao usuário que a palavra "neguinha" possui conotação sexual.
Setembro, 2020	*Historiadoras mulheres e enfermeiros homens não existem para o Google Tradutor*. Estudo demonstrou que o Google Tradutor sistematicamente muda o gênero em traduções quando elas não correspondem a estereótipos.
Maio, 2021	*Facebook esconde postagens sobre o apartheid Israel-Palestina*. Funcionários do Facebook denunciam a aplicação incorreta de regras de moderação que esconde conteúdo de ativistas da Palestina.
Agosto, 2021	*Recurso do Instagram para destacar negócios negros tem efeito contrário*. Criadores negros denunciam que etiqueta sobre "negócios de donos negro/as" derrubou alcance e rendimentos.
Agosto, 2021	*CNDH solicita explicações ao Instagram sobre retirada de conteúdos*. Conselho cobra transparência e informações sobre retirada de conteúdos que tratam de violência policial e racismo.
Setembro, 2021	*Facebook rotula com "primatas" vídeo de homens negros*. IA do Facebook rotulou vídeo de homens negros como "primatas" ao recomendar mais conteúdo.
Dezembro, 2021	*Influenciadores negros recebem 35% a menos.* Estudo nos Estados Unidos demonstrou disparidades significativas em relação a faturamento.
Julho, 2022	*Instagram marca como falso post legítimo sobre branquitude*. Instagram marcou como informação falsa e ocultou *post* de socióloga, por tratar do conceito de branquitude.
Novembro, 2022	*Meta impediu candidatos de impulsionar conteúdo sobre maconha*. Meta contrariou o STF sobre candidatos, mas permitiu a bancos anunciarem sobre o tema.
Novembro, 2022	*YouTube privilegiou Jovem Pan durante eleições*. Estudo identificou priorização de conteúdo do grupo notório por desinformação e posicionamentos antidemocráticos.
Fevereiro, 2023	*Bing Chat defende que "homens brancos cristãos" devem definir o futuro da tecnologia*. Pesquisador expôs como o *chatbot* defendeu repetidamente ideias supremacistas brancas.

Maio, 2023	*Google distribui jogo que simula escravidão para racismo recreativo*. Desenvolvido por produtora brasileira, jogo foi aceito na plataforma de aplicativos *Play Store*.
Junho, 2023	*Sugestão do Google erra nome da própria homenageada*. A página de resultados de busca não reconheceu o nome da homenageada pelo próprio Google.
Junho 2023	*Spotify e Deezer querem que você ouça mais músicas de homens*. Pesquisa apurou que artistas do gênero masculino recebem mais de 55% das recomendações dos algoritmos, contra 20% das mulheres.
Outubro, 2023	*Amazon usou sistemas algorítmicos para estabelecer práticas anti-competição*. Algoritmo foi usado para estimular o encarecimento de diversos produtos, ao forçar competidores a praticarem os mesmos valores da Amazon.
Outubro, 2023	*ChatGPT e Google Bard reproduzem teorias racistas sobre saúde*. Teste identificou que respostas de modelos de linguagem disseminam desinformação racista sobre saúde.
Outubro, 2023	*Meta contribuiu aos abusos contra tigrinos na Etiópia*. Relatório da Anistia Internacional aponta que a Meta contribuiu para violência no conflito.
Dezembro, 2023	*Algoritmo de recomendação da Amazon promove livros com desinformação*. Estudo de 60 mil recomendações identificou a priorização de conteúdo do grupo notório por desinformação e posicionamentos antidemocráticos.
Dezembro, 2023	Twitter libera anúncios nocivos, de supremacia branca a golpes financeiros. Organização CheckMyAds listou exemplos de anúncios ilegais permitidos pela plataforma desde a compra por Elon Musk e a mudança para X.
Dezembro, 2023	*Google e Meta não cumprem promessas de diversidade e inclusão*. Metas e compromissos dos programas de diversidade e inclusão das *big techs* não são cumpridos – corte de até 90% de investimento.
Janeiro, 2024	*MPF investiga conteúdo falso no Kwai*. Inquérito suspeita que a plataforma cria perfis e conteúdos falsos para impulsionar métricas.
Julho, 2024	*Microsoft corta investimentos em diversidade e nega impacto*. Podcast Techish discute como retrocessos na Microsoft impactam tecnologia.
Agosto, 2024	*Instagram falhou em combater discurso de ódio contra mulheres na política dos EUA*. Estudo do CCDH mediu que Instagram falhou em 93% das postagens com discurso de ódio e ameaças.
Setembro, 2024	*Instagram associa termos relacionados a drogas com mulheres negras*. Busca pela palavra “negra” ou derivadas estava bloqueada junto a aviso contra tráfico de drogas.
Outubro, 2024	*Google, Microsoft e Perplexity promovem racismo científico*. Jornalistas identificaram que ferramentas de IA estão usando seu conteúdo, inclusive para gerar resumos e outros tipos de resultados, violando a lei de direitos autorais.
Janeiro, 2014	*Hosts brancos cobram mais por locais equivalentes no Airbnb*. Estudo identificou uma disparidade de 12% em valores que anfitriões brancos e não-brancos conseguem na plataforma.

Fonte: Sistematizada a partir de Silva (2022a).

7.1.3. Aplicativos e Serviços

No período de janeiro de 2010 a outubro de 2024, Silva registrou 62 ocorrências de racismo algorítmico em aplicativos e serviços, incluindo discriminação racial por motoristas de Uber, anfitriões do *Airbnb* e compartilhamento de carros. Em agosto de 2020, policiais novaiorquinos usaram o reconhecimento facial para investigar defensores do movimento do *Black Lives Matter*. Em maio de 2021, um algoritmo em Nova Iorque excluía estudantes negros e latinos das melhores escolas de ensino médio.

Em outubro de 2020, um estudo mostrou que um algoritmo de estimativa de funcionamento dos rins impediu 64 pacientes negros de receber transplantes. Esses casos destacam os prejuízos diretos provocados por vieses algorítmicos em sistemas educacionais e de saúde para minorias étnico-raciais.

Na tabela a seguir, encontra-se o levantamento das ocorrências identificadas por Silva:

Tabela 4 – Mapeamento sobre Aplicativos e Serviços

Data	Ocorrência
Janeiro, 2010	*Câmeras da Nikon não entendem rostos asiáticos*. Recurso para evitar *selfies* com olhos fechados se confunde com olhos de asiáticos.
Março, 2016	*Chatbot da Microsoft torna-se racista em menos de um dia*. A *chatbot* Tay, que constrói discurso a partir de aprendizado de máquina, virou racista e xenófoba em menos de um dia, mostrando falta de compreensão da sociedade pelos engenheiros da empresa.
Agosto, 2016	*PokemonGo reproduziu desigualdades econômicas, marginalizando bairros negros*. Metodologia de distribuição dos "portais" do jogo levou a uma média de 19 portais em bairros negros contra 55 portais em bairros brancos.
Novembro, 2016	*Estudante do MIT denuncia vieses no reconhecimento facial*. Joy Buolamwini realizou TED impactante mostrando como sentiu na pele vieses absurdos de visão computacional: robôs de *startups* não conseguiram reconhecê-la.
Abril, 2017	*App que transforma selfies equipara beleza à brancura*. Aplicativo *FaceApp* viralizou com filtros de vários tipos. O que torna as *selfies* "mais bonitas" também torna os rostos brancos ou mais brancos.
Dezembro, 2017	*Recurso de segurança do sifone confunde rostos de clientes chinesas*. Mãe e filho de Shangai mostram como o recurso falha com suas faces na ferramenta de desbloqueio com reconhecimento facial.
Fevereiro, 2018	*APIs de sistemas de reconhecimento facial não entendem gênero/raça de mulheres negras*. Pesquisadoras Jocy Buolamwini e Timnit Gebro analisaram sistemas da Microsoft, Face++ e IBM e descobriram que a precisão para identificar gênero e idade é muito díspar entre pessoas negras e brancas.
Janeiro, 2019	*Análise facial de emoções associa categorias negativas a atletas negros*. Pesquisadora estudou recurso de análise de expressão facial da Microsoft e Face++ em fotos similares: negros foram marcados como menos Felizes e mais Raivosos.
Junho, 2019	*Perspective classifica tweets de drag queens como "tóxicos"*. *Software* da Alfabeto aplicou índices de "toxicidade" em textos comuns de *drag queens* – maiores do que em textos de supremacistas brancos.
Julho, 2019	*Wearables são mais imprecisos em peles escuras*. Precisão de Fitbit Surge, Samsung Gear e Basis Peak é menor que o aceitável, devido ao tipo de luz usada.
Outubro, 2019	*Escores de risco em saúde penalizam pacientes negros*. Algoritmos de cálculo de necessidade de saúde penalizam pacientes negros, pois se baseiam em indicadores de gastos já racistas.
Abril, 2020	*Aplicativo de transformação de selfie embranquece rostos*. Influenciador MussumAlive e rede mostraram o embranquecimento de suas versões no aplicativo ai-art.tokyo.
Abril, 2020	*Reconhecimento de fala em áudio erra mais com vozes de negros nos EUA*. Estudo identificou taxas de erros discrepantes no reconhecimento de fala em áudio. A métrica de erro médio por palavras (WER) sobe de 0.19 a 0.35 no áudio de afro-americanos.
Junho, 2020	*Aplicativos como Uber e Lyft cobram mais de moradores de bairros periféricos e não-brancos*. Estudo de pesquisadores da George Washington University junto

	ao Censo dos Estados Unidos descobriu que a tarifa dinâmica penaliza com preços relativamente mais altos moradores jovens de bairros com população não-branca.
Agosto, 2020	*GPT-3 associa muçulmanos à violência*. Nova versão de sistema da OpenAI gera com frequência textos associando muçulmanos à violência.
Agosto, 2020	*Corte de imagens no Twitter privilegia rostos brancos*. Algoritmo de cropagem de imagens no *Twitter* privilegia consistentemente rostos brancos.
Setembro, 2020	*Planos de fundo do Zoom escondem professor negro*. Provedor de videoconferência Zoom esconde professor negro ao usar o recurso de plano de fundo virtual.
Outubro, 2020	*Sistema de fotografias para passaporte do Reino Unido é impreciso com mulheres negras*. Usando o Pilot Parliaments Benchmark, a BBC testou o sistema de fotos para o passaporte do Reino Unido e confirmou a disparidade interseccional.
Outubro, 2020	*Algoritmo impede pacientes negros de receber transplante de rim*. Estudo identificou como presunções de diferença racial em algoritmo para estimativa de funcionamento dos rins impediu ao menos 64 pacientes negros de receber o procedimento.
Março, 2021	*Exposição de dados na Holanda permitiu perseguição a minorias*. Órgão governamental perseguiu intencionalmente minorias étnico-raciais no país com o sistema de escrutínio de auxílios-creche.
Março, 2021	*LinkedIn filtra indicações de empregos para imigrantes*. Professor aponta problemas no sistema de recomendação de empregos no *LinkedIn*.
Abril, 2021	*Sistema dá opção de "um pouco de racismo ou xenofobia"*. Intel lançou sistema de filtragem de *chat* por voz que oferece um *slider* ao usuário escolher "Nenhum", "Pouco" ou "Alto" nível de racismo, xenofobia, nazismo e outros tipos de violência discursiva.
Abril ,2021	*YouTube desmonetiza vídeos de Black Lives Matter...* Mas não desmonetiza vídeos de *White Lives Matter*.
Maio, 2021	*Sistema para "prevenção de gravidez" viola direitos de adolescentes do Sul Global*. Microsoft desenvolveu sistema ineficiente que violou direitos de adolescentes na Argentina, no Chile e na Colômbia.
Maio, 2021	*Aplicativo dermatológico foi lançado sem testes apropriados em peles de pessoas negras*. Sem dados ou testes suficientes em pessoas de pele escura, a Google lança aplicativo dermatológico para promover seus negócios no campo da saúde e coletar dados.
Julho, 2021	*TikTok sinaliza "Black Lives Matter" e termos afroamericanos como conteúdo impróprio*. Produtor de conteúdo no TikTok mostrou como a plataforma o impediu de criar conteúdo com *tags* e termos como *Black Lives Matter*, mas permitia sobre supremacismo branco.
Agosto, 2021	*Hipotecas nos EUA são negadas em taxas 90% maiores a pessoas negras*. Metodologia que usa o sistema de escore de crédito Classic FICO penaliza consistentemente pessoas negras, hispânicas e nativas.
Setembro, 2021	*Filtro de sistema de PLN exclui mais documentos ligados a autores negros, hispânicos e LGBTQ+*. Estudo sobre a base de dados Colossal Clean Crawled Corpus identificou como filtros e vieses dos dados penalizaram mais autores negros, hispânicos e LGBTQ+.
Dezembro, 2021	*Aplicativo transfóbico também não reconhece mulheres cis negras*. App Giggle reúne transfobia e supremacismo branco. Possui sistema para impedir mulheres trans de acessar e excluiu também mulheres negras cis.
Janeiro, 2022	*Aplicativo imobiliário aumenta disparidades raciais*. Prometendo justamente o contrário, o aplicativo *Redfin* tem aumentado a escala de discriminação contra propriedades e bairros afro-americanos.
Janeiro, 2022	*Sistema do C6Bank não reconhece correntista negro*. Falhas do sistema são tão vulgares contra correntista negro que até uma foto de ator branco em celular funcionou – mas não a sua.
Março, 2022	*LinkedIn exclui vaga afirmativa de trabalho*. Ignorando constituição brasileira, *LinkedIn* exclui vaga afirmativa de emprego e suspende conta de profissional.

Agosto, 2022	*iPhone não reconhece rostos com Moko Kanohi*. *Software* da Apple não permite o registro de pessoas com tatuagens tradicionais maori.
Novembro, 2022	*Sistemas para contratação discriminam mulheres, idosos e faculdades populares*. Reportagem investiga como serviços como Gupy e Rocketmat podem gerar mais discriminação no mercado de trabalho.
Dezembro, 2022	*Homem processa Apple Watch por não funcionar em peles escuras*. Estadunidense abriu processo federal contra a Apple por desenvolver relógios que não medem corretamente níveis de oxigênio em peles escuras.
Fevereiro, 2023	*Holandesa prova que uso de software a discriminou durante prova*. Estudante holandesa não conseguiu fazer prova on-line da Universidade Vrije, pois o *software* de reconhecimento "anti-trapaça" não a identificou na frente da câmera.
Março, 2023	*Escore de habitação social discrimina pessoas negras sem teto*. Auditoria identificou que sistema VI-SPDAT prejudicou jovens negros sem teto.
Março, 2023	*DALL-E e Stable Diffusion geram representações discriminatórias e estereotipadas*. Pesquisadoras da Universidade Leipzig estudaram 96 mil imagens geradas por modelos de difusão e identificaram tendências discriminatórias de raça, gênero e origem.
Abril, 2023	*Dall-E 2 Mini gera representações de profissões de forma desproporcional*. Sistema generativo de imagens Dall-E superestimou desproporção de gênero e raça, aprofundando estereótipos.
Junho, 2023	*Stable Diffusion exacerba representações negativas ligadas a raça e gênero*. Estudo analisou mais de 5 mil imagens para medir a promoção de representações negativas a pessoas não-brancas.
Junho, 2023	*Recurso do Canva marca estilo de cabelo negro como inseguro*. Ferramenta marcou como "insegura ou ofensiva" a solicitação de geração de imagens de penteado bantu knots.
Agosto, 2023	*Midjourney não consegue gerar imagens de médicos negros tratando crianças brancas*. Experimento de pesquisador sobre representações na saúde prova imagens nocivas sobre pessoas negras.
Agosto, 2023	*Software de detecção de texto gerado por IA errou muito mais com estudantes não-falantes de inglês*. Sistemas como o Turnitin marcaram 61% dos textos escritos por falantes não-nativos como gerados por IA – contra 5,1% dos textos escritos por falantes nativos.
Outubro, 2023	*Clipboard e Shiftkey precarizam trabalho de enfermeira/os*. Empresas desenvolveram algoritmos de *rankings* que precarizam o trabalho de profissionais da enfermagem por aplicativo.
Outubro, 2023	*Midjourney reduz o mundo a estereótipos*. Estudo analisa imagens sobre mulheres, casas, ruas e pratos de comida de diferentes países, identificando redução a estereótipos nas imagens geradas.
Outubro, 2023	*Dall-E inclui armas em prompts sobre mulheres negras em favela*. Deputada Renata Souza tentou gerar uma imagem que correspondesse à sua descrição e foi apresentada a uma ilustração que trazia uma mulher negra segurando uma arma dentro de uma favela.
Novembro, 2023	*Algoritmo de plano de saúde nega 90% dos pedidos*. UnitedHealthcare é processada por supostamente negar 90% de pedidos para pacientes idosos.
Dezembro, 2023	*Base de imagens LAION-RB retirada por conter imagens de abusos contra crianças*. A base de imagens para aprendizado de máquina continha mais de 3 mil imagens suspeitas.
Dezembro, 2023	*Wisconsin usou algoritmo falho de predição de abandono escolar*. Investigação e relato de estudantes prejudicados demonstraram os problemas do sistema, que erra 75% das vezes.
Janeiro, 2024	*Stable Diffusion promove colorismo, sexismo e casteísmo na Índia*. Pesquisador analisou como o modelo gera imagens sobre diferentes categorias, como "Riqueza" ou "Educação", e identificou disparidades.
Janeiro, 2024	*Kwai destaca vídeos que sexualizam crianças*. Mapeamento identificou conteúdos que sexualizam, assediam e chantageiam crianças.

Fevereiro, 2024	*TikTok bane pessoa com nanismo por confundir com criança*. Criadora de conteúdo foi banida da plataforma, pois o sistema a identificou incorretamente como criança.
Março, 2024	*Engenheiro da Microsoft denuncia problemas no DALL-E 3*. Recurso Copilot Designer gerou resultados perturbadores, e o engenheiro denunciou para empresa, imprensa e órgão regulador.
Maio, 2024	*Canva gera apenas garotos negros em prompt ligado a tornozeleiras eletrônicas*. Pesquisadora demonstra como *prompt* sobre "garotos usando tornozeleira" gerou apenas imagens de crianças negras.
Julho, 2024	*LAION-5B usou imagens de crianças sem permissão*. Estudo australiano identificou que a base de dados utilizou imagens de crianças sem conhecimento ou autorização.
Julho, 2024	*Recurso do Canva inclui cocaína em foto de intelectual negra*. O recurso de expandir imagens do Canva inclui cocaína em foto de Lélia González, denuncia jornalista.
Julho, 2024	*Seguradoras de Michigan cobram mais de motoristas negros*. Estudo analisou como algoritmos de precificação para seguros discriminam bairros predominantemente negros.
Julho, 2024	*Algoritmo VioGén prevê baixo risco de feminicídio – e erra fatalmente*. Sistema para projetar risco de violência doméstica descarta proteção a mulher, assassinada em seguida. Outros 50 casos foram identificados.
Julho, 2024	*Trabalhadores relatam queda na produtividade e aumento na sobrecarga graças a I.A.* Estudo da UpWork mediu que 77% dos trabalhadores foram prejudicados pelas iniciativas corporativas de implementação de IA.
Agosto, 2024	*Llama3 e GPT4 falham em sumarização de informações médicas*. Pesquisadores denunciam que 40% dos resumos incluíram informações erradas, e todos apresentaram hipergeneralização.
Setembro, 2024	*Wikipedia cria grupo para limpar as páginas de desinformação gerada com IA*. Multiplicação da desinformação afeta qualidade da enciclopédia aberta.
Outubro, 2024.	*Whisper inventa coisas em transcrição para saúde*. Serviço de transcrição da OpenAI usado em hospitais inventa coisas que ninguém disse.

Fonte: Sistematizada a partir de Silva (2022a).

7.1.4 Tecnologias de Vigilância e Ordenação

No período de junho de 2015 a junho de 2024, Silva registrou 36 ocorrências de racismo algorítmico relacionadas ao uso de tecnologias de reconhecimento facial para vigilância e ordenação de cidadãos, com casos de erros e discriminação racial. Esses sistemas penalizaram grupos marginalizados na justiça e no mercado financeiro. Estudos de agosto de 2020 mostraram que algoritmos de reconhecimento facial discriminam afro-americanos e asiáticos mais do que caucasianos.

Em novembro de 2019, dados mostraram que 90,5% das detenções no Brasil baseadas em reconhecimento facial envolviam negros. Em junho de 2024, o documentário *Mind the People* revelou como sistemas algorítmicos discriminam comunidades na África do Sul no acesso a serviços públicos.

A sistematização dessas ocorrências se encontra na tabela a seguir:

Tabela 5 – Mapeamento de Tecnologias de Vigilância e Ordenação

Data	Ocorrência
Junho, 2015	*Terrorista foi racializado por resultados no Google*. Terrorista que assassinou 9 pessoas foi racializado por resultados do Google – que sugerem termos errôneos sobre crimes nos Estados Unidos.
Maio, 2016	*Startup israelense alega identificar traços faciais de terroristas*. Faception alega identificar características faciais de terroristas e pedófilos.
Maio, 2016	*Software de análise de reincidência prejudica réus negros e favorece réus brancos*. Projeto da ProPublica analisou o *software* COMPAS[20] e denunciou como "errou" ajudando réus brancos e prejudicando réus negros.
Outubro, 2018	*Amazon vende reconhecimento facial para o ICE*. Amazon vende reconhecimento facial para o ICE, órgão que reprime migração ilegal, responsável pelos campos de concentração nos Estados Unidos.
Março, 2019	*Carros autônomos têm mais chance de atropelar pessoas negras*. Pesquisadores da George Institute of Technology descobriram que a visão computacional em sistemas de carros autônomos foi treinada para identificar melhor pedestres de pele clara.
Abril, 2019	*China acusada de perfilar minoria muçulmana com biometria*. Governo chinês está sendo acusado de usar IA para identificar a minoria étnico-religiosa Uighur.
Setembro, 2019	*Algoritmo de identificação de discurso de ódio discrimina linguagem afro-americana*. Teste com o sistema Google Perspective, usado para moderar conteúdo on-line, mostrou discriminação contra linguagem afro-americana.
Setembro, 2019	*Recurso de foto para passaporte confunde lábios de britânico*. Sistema de fotografia para passaporte britânico pediu a jovem negro que "fechasse a boca" ao não entender seus lábios.
Novembro, 2019	*Erros de reconhecimento facial da polícia baiana traumatizam jovens negros*. A polícia da Bahia coleciona erros de abordagem a partir de reconhecimento facial de suspeitos e foragidos.
Novembro, 2019	*90,5% dos presos por reconhecimento facial no Brasil são negros*. Rede de Observatórios da Segurança lançou dados sobre as prisões baseadas em reconhecimento facial, Bahia lidera o número de abordagens e prisões.
Dezembro, 2019	*Agência estadunidense publica estudo sobre erros racistas no reconhecimento facial*. Algoritmos identificam faces afro-americanas e asiáticas erroneamente de 10 a 100 vezes em comparação a faces caucasianas.
Agosto, 2020	*Polícia de Nova Iorque usa reconhecimento facial para perseguir ativista*. Rede de lojas implementou 3 vezes mais câmeras de reconhecimento facial em bairros negros/latinos. Rede de farmácias RiteAid implementou reconhecimento facial de forma desproporcional em bairros com população negra e/ou latina.
Agosto, 2020	*Algoritmo para vistos do Reino Unido discrimina por raça e região*. Grupo do Reino Unido conseguiu a suspensão do uso de algoritmo.
Agosto, 2020	*Polícia de Nova Iorque usa reconhecimento facial para perseguir ativista*. A polícia de Nova Iorque utilizou reconhecimento facial para perseguir ativista do *Black Lives Matter* que não cometeu crime.
Fevereiro, 2021	*Assistentes digitais não compreendem adequadamente sotaques*. Estudos mostraram que assistentes como Siri e Alexa não compreendem adequadamente sotaques, como o AAVE (african-american vernacular english) e o de migrantes.
Abril, 2021	*Startup Unico IDTech obriga pessoas vulneráveis a fornecer dados biométricos*. Em parceria com ONG, *startup* de biometria, identificação e vigilância vincula ação social ao fornecimento de dados biométricos como rosto para reconhecimento facial.

[20] COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) ou Perfil de Gestão de Infratores Correcionais para Sanções Alternativas é uma ferramenta de gestão de casos e suporte à decisão desenvolvida e de propriedade da Northpointe, usada pelos tribunais dos EUA para avaliar a probabilidade de um réu se tornar um reincidente. Trata-se de um sistema usado em estágios sequenciais da justiça criminal, incluindo: correções, pré-julgamento, liberdade condicional, detenção e prisões. Disponível em: https://www.researchgate.net/publication/321528262_Correctional_Offender_Management_Profiles_for_Alternative_Sanctions_COMPAS. Acessado em 10/06/2024.

Julho, 2021	*Sensores automáticos de tiroteios penalizam bairros negros*. Uso do ShotSpotter, sistema de sensores de som que gera alertas de possíveis tiroteios, foi desproporcionalmente implementado em bairros negros e gera alertas falsos frequentemente.
Setembro, 2021	*Instituto Igarapé tenta expandir projeto de hipervigilância*. Projeto CrimeRadar promove hipervigilância e violência policial com a falsa ideia de "policiamento preditivo".
Outubro, 2021	*Uber exclui erroneamente motoristas com reconhecimento facial falho*. Sindicato na Grã-Bretanha acusa Uber de usar sistema de reconhecimento facial que excluiu injustamente motoristas de minorias étnico-raciais.
Janeiro, 2022	*Câmeras automáticas de trânsito penalizam bairros negros e latinos em Chicago*. Fardo e impacto das multas aumentam disparidades que estão ligadas a padrões urbanísticos, apontam pesquisadores.
Janeiro, 2022	*Algoritmo Pattern de reincidência criminal penaliza presos negros, latinos e asiáticos*. Sistema de predição de risco de reincidência criminal nega liberdade condicional de forma desproporcional.
Agosto, 2022	*Stable Diffusion associa gangues a homens negros*. Modelo de geração de imagens artificiais apresenta só fotos de homens negros em reação a termos sobre gangues.
Setembro, 2022	*Amsterdam cria listas preditivas para perseguir jovens*. Mãe relata algoritmização da seletividade penal em listagens da polícia de Amsterdam.
Novembro, 2022	*Prefeitura de SP lança edital racista para reconhecimento facial*. A prefeitura de São Paulo lançou editou sobre reconhecimento facial, incluindo até categorias como "vadiagem".
Dezembro, 2022	*Deputada quer barrar reconhecimento facial em SP por racismo*. Deputada Erika Hilton age contra proposta racista e vigilantista do Smart Sampa.
Fevereiro, 2023	*Nova Iorque priorizou bairros negros para vigilância biométrica*. Relatório da Anistia Internacional demonstrou que a distribuição de câmeras de reconhecimento facial na cidade penalizou bairros com população não-branca.
Março, 2023	*Cidade de Rotterdam usa software discriminatório na identificação de fraudes*. *Software* desenvolvido com a Accenture acusa desproporcionalmente trabalhadores migrantes.
Julho, 2023	*Algoritmo de identificação de fraude foi usado para perseguir estudantes de minorias*. Organização DUO da Holanda aplicou algoritmo que acusou falsamente de fraude estudantes de grupos minoritários.
Setembro, 2023	*Refugiados são prejudicados por softwares de tradução automática nos EUA*. Refugiado brasileiro foi prejudicado por *software* que não reconhecia sequer o nome da cidade Belo Horizonte.
Outubro, 2023	*Instagram inclui palavra "terrorista" em bios de palestinos*. Tradução usada pela Meta transformou várias combinações de termos em "terrorista" nas traduções.
Outubro, 2023	*Contas pró-Palestina bloqueadas em mídias sociais*. Perfis informativos do ponto de vista palestino foram bloqueados em mídias como Instagram por supostas questões de segurança.
Dezembro, 2023	*Rede de farmácia é proibida de usar reconhecimento facial*. Após denúncia de violações de direitos, órgão estadunidense proíbe Rite Aid de implementar reconhecimento facial por 5 anos.
Janeiro, 2024	*Pessoas presas no Rio com reconhecimento facial não tinham mandados*. Matéria denuncia casos de erros ligados aos bancos de dados de mandados no sistema implementado no Rio de Janeiro.
Maio, 2024	*Clientes acusados de serem ladrões pelo reconhecimento facial*. Sistemas da Facewatch constrangem consumidores em redes de loja no Reino Unido.
Junho, 2024	*Engenheiro da Meta processa a empresa por discriminação ligada a Gaza*. Engenheiro contou que foi demitido depois de buscar ajustar *bugs* que prejudicavam *posts* de pessoas da Palestina.

Fonte: Sistematizada a partir de Silva (2022a).

7.1.5 Impactos Sociais e Políticos

Entre agosto de 2016 e outubro de 2024, foram registradas 20 ocorrências de racismo algorítmico com impactos sociais e políticos. Esses casos mostram como essa prática afeta profundamente a vida de minorias étnico-raciais. Tecnologias digitais influenciaram eleições e movimentos sociais, como o *Brexit* e as eleições americanas, e grandes empresas de tecnologia fizeram *lobby* para influenciar políticas públicas.

Em 2016, um projeto de aprendizado de máquina em concursos de beleza foi criticado por padrões racistas. Em 2017, o Facebook foi acusado de expor vítimas de violência doméstica ao exigir nomes reais. No ano de 2018, uma pesquisa revelou que mulheres negras e indígenas eram sub-representadas na área de tecnologia no Brasil. Em 2020, o Facebook foi criticado por ignorar estudos sobre vieses raciais e perfis falsos que simulavam afro-americanos apoiando figuras políticas controversas como Donald Trump, distorcendo a realidade e influenciando a opinião pública.

Esses casos levantamos na *Timeline* de Silva foram sistematizados e se encontram na tabela a seguir:

Tabela 6 – Mapeamento de Impactos Sociais e Políticos

Data	Ocorrência
Agosto, 2016	*Projeto aplica aprendizado de máquina a concurso de beleza*. Projeto apoiado por empresas de cosméticos reproduz e vulgariza padrões racistas de beleza ao aplicar aprendizado de máquina.
Abril, 2017	*Facebook expôs vítimas de violência doméstica*. Política do Facebook em exigir nomes reais expôs usuários.
Janeiro, 2018	*Pesquisa mapeia mulheres negras e indígenas na tecnologia no Brasil*. Pesquisa inédita da PretaLab levantou dados sobre áreas de atuação, perfil de formação e contato com a área.
Julho, 2020	Facebook ignorou e impediu estudos internos sobre vieses raciais na moderação automatizada de conteúdo no Instagram.
Agosto, 2020	*Perfis simulam afro-americanos apoiando Trump e extremistas*. Contas inautênticas simulavam ser afro-americanos.
Outubro, 2020	*Instagrammer negra compara entrega algorítmica de conteúdo na plataforma*. Sá Ollebar realizou experimento de comparação de entrega algorítmica de conteúdo na plataforma: fotos de mulheres brancas em seu próprio perfil ganharam mais destaque.
Outubro, 2021	*Vídeos transfóbicos no TikTok levam também a mais conteúdos racistas e misóginos*. Experimento sobre transfobia e algoritmos no TikTok mostrou como o sistema recompensa conteúdo extremista.
Dezembro, 2021	*Refugiados Rohingya buscam reparação pelo genocídio*. Refugiados rohingya buscam 150 bilhões de reparação pelo papel do Facebook em genocídio.
Julho, 2022	*Startup usa IA para eliminar "sotaques"*. Empresa lança *software* para transformar sotaques diversos no padrão branco americano, mirando terceirização de teleatendimento.
Setembro, 2022	*Desenvolvedor substitui atriz negra por atriz ruiva em trailer*. Usando *deepfake*, programador substitui a atriz Chloe Bailey de trailer de *A Pequena Sereia*.

Agosto, 2023	*Scale AI precariza fornecedores falantes de línguas do Sul Global*. Sistema de anotação para aprendizado de máquina paga até 15 vezes menos para treinamento em línguas de países do Sul Global.
Setembro, 2023	*Impacto da seca em Iowa intensificado pela Microsoft*. Sistemas de resfriamento para IA generativa teriam intensificado impactos da seca em Iowa.
Novembro, 2023	*Ex-Chefe de Segurança no Twitter denuncia plataforma*. Chefe de Confiança e Segurança no Twitter contou sobre postura que priorizou escala e crescimento contra segurança.
Novembro, 2023	*I.A. produzida para facilitar assassinatos em massa em Israel*. Reportagem investiga como o governo israelense aplicou IA para facilitar ofensas permissivas no genocídio.
Dezembro, 2023	*Gerar uma imagem por I.A. gasta tanto quanto carregar um smartphone*. Estudo analisou impacto energético de serviços de produção de imagens por IA derivativa.
Dezembro, 2023	*Meta permitiu anúncios contra Palestina*. Anúncios que promoveram a teoria da conspiração *Pallywood* foram permitidos na plataforma.
Março, 2024	*ABL reforça falsa branquitude de Machado de Assis com IA*. Academia Brasileira de Letras lançou totem. A Academia Brasileira de Letras usou a Inteligência Artificial para trazer o escritor Machado de Assis para o século XXI. A figura ilustre do escritor, falecido em 1908, passou a recepcionar os visitantes.
Junho, 2024	*Twitter / X monetizou termos e hashtags racistas*. Novas evidências sobre como o Twitter, atual X, gerou receita em páginas ligadas a temos como #whitepower.
Agosto, 2024	*Modelos algorítmicos de linguagem discriminam dialetos afro-americanos*. Estudo identifica que modelos de linguagem associa dialetos afro-americanos a valores negativos.
Setembro, 2024	*Estudantes negros são mais acusados de usar IA generativa*. Trabalho identificou que estudantes negros reportam mais acusações de terem realizado trabalhos com IA generativa.

Fonte: Sistematizada a partir de Silva (2022a).

A forma de documentação das ocorrências de casos apresentados por Tarcízio Silva na *Timeline* descreve os eventos, inclui dados, reações da sociedade civil, respostas das empresas envolvidas e, em certos contextos, referências a trabalhos acadêmicos ou reportagens que possibilitam, neste estudo, um aprofundamento na análise. Essa contextualização é fundamental para uma pesquisa que se propõe a analisar o racismo algorítmico à luz do biopoder de Foucault, pois permite compreender não apenas os eventos em si, mas também os discursos e as práticas que os constituem, atualizam e perpetuam.

A abordagem teórica da pesquisa, baseada no conceito de biopoder de Foucault, demanda uma análise dos discursos e das práticas que atualizam, fortalecem e reproduzem o racismo algorítmico como uma forma de controle social e racialização. A Linha do Tempo do Racismo Algorítmico, ao documentar não apenas os casos, mas também as reações e os debates que eles geram, oferece um material significativo para essa análise, pois permite identificar como os discursos sobre racismo e tecnologia são construídos, contestados e transformados ao longo do tempo.

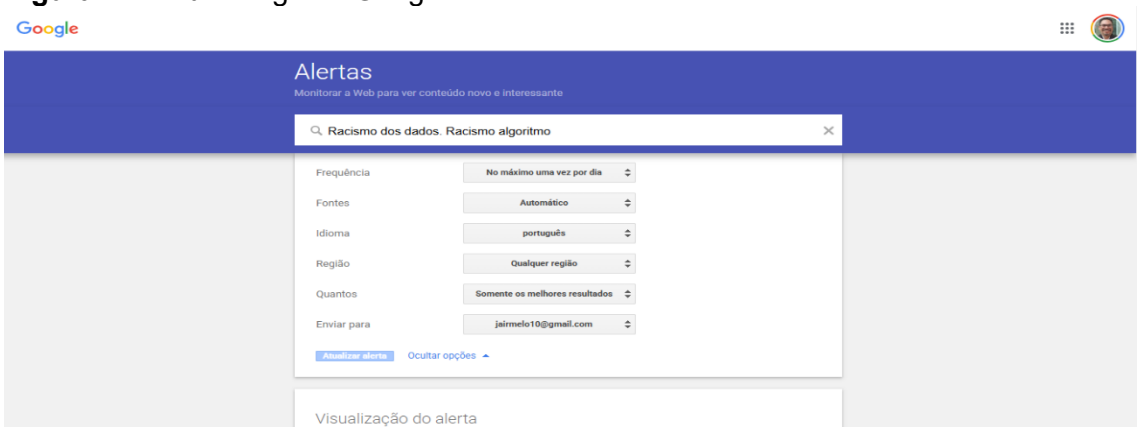
Além disso, a abrangência temporal da *Timeline* permite identificar não apenas casos isolados, mas também padrões e tendências ao longo dos anos, algo essencial para uma análise foucaultiana do biopoder, que se preocupa com a forma como o poder se manifesta e se transforma na sucessão da linha temporal.

A opção pela *Timeline* como fonte de pesquisa não excluiu a possibilidade de complementar o *corpus* com outras fontes, como notícias recentes, artigos acadêmicos ou dados coletados por meio de ferramentas como o Google Alerts, com o objetivo de expandir e atualizar o estudo, garantindo a abrangência e a contemporaneidade desta pesquisa, a qual não apenas descreve, mas também problematiza e contribui para o debate acadêmico e público sobre justiça algorítmica e desigualdades raciais.

7.2 Seleção das Ocorrências de Racismo Algorítmico do Google Alerts

Como segundo eixo, para complementar o *corpus* constituído pela *Timeline* abordada anteriormente, foram utilizadas notificações da ferramenta Google Alerts, configuradas para monitorar termos como “racismo algorítmico”, “viés racial em algoritmos”, “reconhecimento facial e raça”, entre outros relacionados, no período de 22/06/2022 a 15/12/2024, tal como se pode visualizar na figura a seguir:

Figura 1: *Print* de registro Google Alerts.



Registro da configuração do alerta criado em 12/06/2022.

Essa ferramenta permitiu identificar casos recentes e debates emergentes sobre o tema, ampliando o escopo da pesquisa. A ferramenta gratuita, oferecida pelo Google, permite que se monitore o ciberespaço a partir de um comando configurado

pelo usuário. Este, pela inserção de uma palavra-chave, frase ou tópico de interesse, cria um “alerta” e recebe notificações por e-mail toda vez que algum conteúdo correspondente aos termos selecionados é publicado. Essa funcionalidade, para efeito de constituição do *corpus* desta pesquisa, foi ativada em 12/06/2022, para monitorar e filtrar menções sobre o tema do trabalho.

A escolha do Google Alerts como ferramenta complementar para a constituição do *corpus* da pesquisa sobre o tema do racismo algorítmico à luz do biopoder de Foucault justifica-se por questões metodológicas, as quais estão alinhadas tanto às especificidades do objeto de estudo quanto aos objetivos da investigação.

Como uma ferramenta que atual na diversidade dos espaços virtuais, repletos meio de notícias, artigos acadêmicos, blogs, fóruns de discussão e redes sociais, o Google Alerts, nesta pesquisa, possibilitou monitorar as ocorrências de racista algorítmico em variadas fontes indexadas pelo Google, o que é essencial para capturar a multiplicidade de discursos e práticas relacionadas ao tema. Essa amplitude da ferramenta de busca é significativa para uma pesquisa que se propõe a compreender como esses discursos circulantes se inscrevem nos espaços virtuais, como se reconfiguram e como são discutidos na sociedade como um todo. Para Foucault (1998, 2008), é necessário analisar os discursos como práticas que produzem realidades e sujeitos. Ao monitorar as ocorrências de racismo algorítmico por meio do Google Alerts, esta pesquisa conseguiu mapear os discursos que constituem, atualizam e solidificam esse modelo de racismo.

Outro aspecto relevante da ferramenta de busca do Google, neste estudo, refere-se ao dispositivo de receber notificações, que opera em tempo real, o que possibilita que o *corpus* da pesquisa esteja atualizado, com ocorrências mais recentes sobre o tema, diante de um fenômeno marcado pela dinâmica da evolução e da transformação tecnológica e social. Cabe salientar que esse mecanismo, adotado como complementar à constituição do *corpus*, apresenta limitações, uma vez exclui conteúdos não indexados pelo Google. No entanto, essas limitações podem ser contornadas por meio da complementaridade com outras metodologias, como análise de redes sociais, revisão bibliográfica sistemática ou entrevistas com especialistas, que podem ser levadas a cabo em outras pesquisas.

No período delimitado para a pesquisa a partir da configuração do Google Alerts – 22/06/2022 a 15/12/2024 –, para constituição complementar do *corpus* desta

pesquisa, foram observados 101 alertas de textos, artigos, posts, identificados por meio dessa ferramenta de monitoramento.

A partir da seleção do arquivo digital consultado, realizou-se a organização dos dados em três eixos de ocorrência, a saber:

a) Reconhecimento facial aplicado à segurança pública e à justiça criminal;

b) Testes padronizados usados na saúde e sistema de pontuação de crédito e seleção de emprego; e

c) Sistemas de branqueamento em filtros de aplicativos, Imagens geradas por inteligência artificial (IA) e sistemas de recomendação usados nas redes sociais.

Esses itens foram tabulados e encontram-se sistematizados na tabela a seguir:

Tabela 7 – Grade da Seleção das Ocorrências de Racismo Algorítmico do Google Alerts

Campo de ocorrência	**Data**	**Fonte**	**Exemplo**
a) Reconhecimento facial aplicado à segurança pública e justiça criminal	09/08/2023, 14h10	www.poder360.com.br	Título: *Futuro Indicativo: Prefeito de SP vende reconhecimento facial como solução mágica*. Link: https://www.poder360.com.br/futuro-indicativo/prefeito-de-sp-vende-reconhecimento-facial-como-solucao-magica/
	11/04/2023, 14h28	www.noticiapreta.com.br/contato/,by Mariane Del Rey	Título: *Inocentado anteriormente, homem é acusado de roubo de novo por reconhecimento facial*. Link: https://noticiapreta.com.br/homem-negro-roubo-econheciemto-facial/
b) Testes padronizados usados na saúde; sistema de pontuação de crédito e seleção de emprego	25/10/2022, 13h16	www.science.org, por Ziad Obermeyer	Título: *Dissecando o preconceito racial em um algoritmo usado para gerenciar a saúde das populações*. Link: https://www.science.org/doi/10.1126/science.aax2342
	26/10/2022, 15h23	www.wired.com, por Tom Simonite	Título: *Como um algoritmo bloqueou transplantes renais para pacientes negros*. Link: https://www.wired.com/story/how-algorithm-blocked-kidney-transplants-black-patients/
c) Sistemas de branqueamento em filtros de aplicativos; Imagens geradas por inteligência artificial (IA) e sistemas de recomendação usados nas redes sociais	20/04/2023, 14h32	*New York Times*, por Zachary Small	Título: *O que essa tecnologia está fazendo com a história? Artistas negras apontam viés racista em inteligência artificial*. Link: https://oglobo.globo.com/cultura/noticia/2023/07/09/o-que-essa-tecnologia-esta-fazendo-com-a-historia-artistas-negras-apontam-vies-racista-em-inteligencia-artificial.ghtml
	23/07/2023, 14h07	*Jornal do Brasil*, com alma preta jornalismo redacao @jb.com.br	Título: *Feia x bonita: como o racismo algorítmico impacta imagem de mulheres negras na internet*. Link: https://www.jb.com.br/brasil/2023/07/1045052-feia-x-bonita-como-o-racismo-algoritmico-impacta-imagem-de-mulheres-negras-na-internet.html.

	06/10/2023, 07h44	*Goats and Soda: stories of life in a changing world*, by Carmen Drahl	Título: *Foi pedido à AI que criasse imagens de médicos negros africanos tratando crianças brancas*. *Como foi?* Link: https://www.npr.org/sections/goatsandsoda/2023/10/06/1201840678/ai-was-asked-to-create-images-of-black-african-docs-treating-white-kids-howd-it-
	13/11/2023, 11h53	www.terra.com.br/byte/	Título: *Deputada do Rio denuncia à polícia racismo de imagem gerada por IA*. Link: https://www.terra.com.br/byte/deputada-do-rio-denuncia-a-policia-racismo-de-imagem-gerada-por-ia,c8796bf8d62ac9ebe0d06a544e283813cgzn6eyp.html
	22/11/2023, 14h06	*Le Monde,* caderno Diplomatique Brasil	Título: *Racismo algorítmico: a exclusão da população negra*. Link: https://diplomatique.org.br/racismo-algoritmico/
	28/04/2024, 11h59	*El País*, Madri-25 Abr. 2017	Título: *Aplicativo FaceApp "branqueia" os usuários para torná-los "mais sexy"*. Link: https://brasil.elpais.com/brasil/2017/04/25/tecnologia/1493122888_029183.html

Fonte: Criação própria, a partir de dados levantados do Google Alerts.

7.3 Procedimentos de Análise

A análise do *corpus* foi realizada a partir dos seguintes procedimentos: os casos e os exemplos foram selecionados com base em sua relevância para o tema do racismo algorítmico e sua conexão com o conceito foucaultiano de biopoder. Os dados foram organizados em três grandes eixos temáticas, os quais agrupam algumas das áreas mais comuns de denúncia do racismo algorítmico e foram elencados no tópico anterior: a) Reconhecimento facial aplicado à segurança pública e justiça criminal; b) Testes padronizados usados na saúde; Sistema de pontuação de crédito e seleção de emprego; c) Sistemas de branqueamento em filtros de aplicativos; imagens geradas por inteligência artificial (IA) e sistemas de recomendação usados nas redes sociais.

Trata-se de uma abordagem essencialmente qualitativa, focada na compreensão dos processos que estruturam os discursos racistas, e não em resultados que possam ser quantificados ou analisados empiricamente. Além disso, ao adentrar o universo teórico de Foucault, a arqueologia se torna uma ferramenta metodológica necessária para essa investigação. Dessa forma, parte-se da compreensão do racismo algorítmico como um discurso, uma prática discursiva que, ao ser analisada teoricamente, revela uma arqueologia que o constitui como objeto e

veículo de enunciados, identificando-se as regras que determinam a emergência de conceitos e suas transformações em curso, além de posicioná-la como a teoria e o tema estruturados dentro de estratégias de poder.

A análise das representações imagéticas é conduzida a partir das perspectivas da genealogia e da arqueologia propostas por Foucault, para compreendê-las a partir de efeitos de sentido fluidos, não rigidamente atrelados a uma linearidade histórica. Seu caráter dispersivo delimita posições sociais, funcionando como um ponto de enunciação para sujeitos em diferentes contextos.

Desse modo, as duas abordagens metodológicas são complementares e permitem compreender as estratégias discursivas usadas pelos sujeitos que se deslocam e se redefinem historicamente. Ao explorar a arqueologia, identificam-se vestígios os quais evidenciam como a representação do negro, relativamente à sua imagem, vem se consolidando, ao longo do tempo, nos ambientes virtuais, como uma prática que sustenta o branqueamento, reforçando sua regularidade no imaginário social.

Nesse contexto, não se objetiva rastrear uma origem definitiva do viés algorítmico, mas identificar as relações de poder e os jogos de força que sustentam esses discursos discriminatórios e que contribuem para a construção e a manutenção de um padrão de branquitude, a qual opera como um dispositivo de poder, influenciando identidades e subjetividades hodiernamente. Compreende-se como os discursos se inserem em um ambiente social permeado por relações de força, que conferem poder a determinados sujeitos ao longo do tempo e que permitem que discursos contemporâneos reapareçam, mesmo que pareçam esquecidos ou adormecidos.

Relativamente ao discurso e à sua estrutura, nas palavras de Foucault,

> Pode-se, creio eu, isolar outro grupo de procedimentos. Procedimentos internos, visto que são os discursos eles mesmos que exercem seu próprio controle; procedimentos que funcionam, sobretudo, a título de princípios de classificação, de ordenação, de distribuição, como se se tratasse, desta vez, de submeter outra dimensão do discurso: a do acontecimento e do acaso (2013, p.20).

É pertinente, para o decurso deste estudo, trazer como adendo a definição de "comentário"[21], por ser complementar no processo analítico em questão. Trata-se do primeiro procedimento interno ao discurso, segundo Foucault (2013), cujo princípio afirma que, dada a raridade dos discursos, muitos daqueles que circulam, em verdade, são formas repetíveis de discursos já existentes, ou seja, são discursos novos, porém sem novidade: "O novo não está no que é dito, mas no acontecimento de sua volta" (FOUCAULT, 2013, p.26).

Assim sendo, comentário é um processo que evidencia um desnível entre diferentes tipos de discursos: de um lado, há aqueles que surgem e se dissipam no fluxo cotidiano das interações, existindo apenas quando são enunciados, e, de outro, há os discursos que continuam a ecoar, servindo de base para novas falas que os retomam, modificam ou reinterpretam. Esses últimos não se limitam ao instante da formulação inicial, mas permanecem vivos e constantemente reatualizados, sendo retomados e reformulados ao longo do tempo (FOUCAULT, 2013, p.21).

Trata-se de uma observação acerca da permanência de certos discursos em perspectiva histórica, bem como da impermanência de algumas de suas referências ou formas. Isso significa que o deslocamento que constitui o comentário não é estável, tampouco absoluto. Nesse sentido, conforme Foucault, "Muitos textos maiores se confundem e desaparecem, e, por vezes, comentários vêm tomar o primeiro lugar" (2013, p.23). Assim, o comentário é o princípio interno que permite a classificação e a categorização dos discursos, dada sua repetição em distintas materialidades históricas. Portanto, existe um paradoxo constante no qual se deve afirmar algo como se fosse a primeira vez, mesmo que esse enunciado já tenha sido expresso anteriormente. Ao mesmo tempo, é necessário repetir incessantemente aquilo que, paradoxalmente, nunca foi dito antes. Esse ciclo de repetição infinita dos comentários carrega em si o desejo de uma repetição transformada, como se, no fundo, tudo que se projeta no futuro já estivesse presente desde o início, reduzindo-se à mera recitação (FOUCAULT, 2013, p.25).

[21] Foucault (2013) define o comentário como um mecanismo fundamental dentro dos discursos, funcionando como um processo de retomada e transformação do que já foi dito. O autor distingue dois tipos de discursos: aqueles que surgem no cotidiano e desaparecem com o ato de fala e aqueles que permanecem, sendo constantemente retomados, modificados e reinterpretados. O comentário não apenas repete um discurso anterior, mas também o reorganiza, atribuindo-lhe novos sentidos e permitindo sua continuidade ao longo do tempo. Esse processo de repetição e transformação é essencial para a circulação dos discursos, pois possibilita que ideias sejam constantemente ressignificadas dentro de diferentes contextos sociais e históricos.

O princípio do comentário, sobretudo, sinaliza a rarefação dos discursos através do esforço de compilação de distintas versões em que o mesmo material emergiu investido de ares de novidade, atuando como um indicador da raridade dos discursos, revelando que, quando um sujeito se expressa, frequentemente percorre trajetórias já estabelecidas. No entanto, busca inovar, ora se submetendo às forças que moldam identidades, ora desafiando essas imposições, ou ainda se posicionando como sujeito que transcende tanto a submissão quanto a resistência, apresentando seu discurso como algo novo. Dessa maneira, os processos de subjetivação e objetivação se manifestam de forma contínua, sempre em movimento, impulsionados pela dinâmica do poder.

Sendo assim, a utilização do conceito de comentário se mostra pertinente em face da dinâmica encontrada nas manifestações do racismo algorítmico, pois elas, embora se apresentem travestidas em algo novo, são, na verdade, a retomada de discursos já ditos em outros tempos.

7.3.1 Reconhecimento facial aplicado à segurança pública e à justiça criminal

As novas tecnologias digitais não apenas refletem as normas e os valores existentes, mas também têm o potencial de remodelar normas por intermédio de sua operação, reforçando ou desafiando padrões existentes de poder e privilégio. Inúmeros são os casos de uso de algoritmos em *softwares* de reconhecimento facial (SILVA, 2022A; SARLET, 2022; BENJAMIN, 2021) com expressivos percentuais de erros significativamente maiores para rostos de sujeitos negros em comparação com rostos de brancos. Essa disparidade ocorre não devido a uma programação intencional, mas por uma falha em fornecer um conjunto de dados diversificado e representativo, mas também pode ocorrer devido a preconceitos históricos na coleta de dados ou a limitações tecnológicas que favorecem certos grupos.

Nesse contexto, Bezerra *et al.* (2022, p.5) afirmam que a intersecção do racismo algorítmico com os sistemas de justiça criminal e vigilância representa uma das áreas mais críticas e preocupantes no estudo das dinâmicas sociais contemporâneas. A crescente dependência de algoritmos para informar decisões judiciais, proceder monitoramento policial e desenvolver práticas de vigilância tem revelado um padrão perturbador: a propensão dessas tecnologias para intensificar e, em alguns casos, perpetuar a discriminação racial, cujas consequências de

preconceitos incorporados podem ser particularmente severas, afetando a liberdade e a vida de grupos racializados.

A implementação de algoritmos na justiça criminal, desde a previsão de riscos de reincidência até o direcionamento de patrulhas policiais, tem sido apontada como uma solução para tornar o sistema mais eficiente e imparcial. Contudo, estudos têm demonstrado que esses sistemas podem codificar e reforçar preconceitos raciais existentes na sociedade. Isso ocorre devido, em parte, a conjuntos de dados históricos utilizados para treinar tais algoritmos, que refletem disparidades raciais em taxas de criminalidade, resultando em previsões que desproporcionalmente marcam certas etnias como sendo de maior risco.

Para Coimbra *et al.* (2023, pp.18-19), a precisão dessas tecnologias frequentemente varia entre diferentes grupos raciais, com taxas significativamente mais altas de falsos positivos para minorias raciais. Essa imprecisão não é apenas um problema técnico, mas uma questão de justiça social, pois aumenta as chances de sujeitos inocentes serem submetidos a investigações, detenções e até violência policial baseadas em identificações incorretas. Assim, o impacto do racismo algorítmico na justiça criminal e na vigilância não é uma consequência inevitável da tecnologia, mas o resultado de escolhas feitas em seu desenvolvimento e em sua implementação.

No sistema de justiça criminal, algoritmos de predição de risco, como o COMPAS, comentado em nota anterior, têm sido amplamente utilizados para determinar fianças, sentenças e decisões de liberdade condicional. A dependência dessas ferramentas tecnológicas pode resultar em decisões enviesadas que perpetuam discriminações raciais e sociais. Esses exemplos ilustram como sistemas automatizados podem refletir e amplificar desigualdades sociais preexistentes, como no caso da aplicação do *software* COMPAS, utilizado nos Estados Unidos, em 2016, para avaliar a probabilidade de reincidência criminal, em que o algoritmo atribuía uma maior probabilidade de reincidência a réus negros em comparação com réus brancos, mesmo quando os fatores de risco eram semelhantes. Essa disparidade resultou de dados históricos enviesados que alimentaram o sistema, reforçando preconceitos raciais que ainda perduram até os dias atuais.

Buolamwini & Gebru (2018), ao examinarem programas de reconhecimento facial nos Estados Unidos, constataram que os algoritmos de análise facial, os quais utilizam *machine learning*, frequentemente eram treinados em bases de dados

desequilibradas em relação à raça e ao gênero. Isso significa que esses sistemas podem ter dificuldade em reconhecer com precisão rostos pertencentes a grupos sociais específicos. O sistema de reconhecimento facial da IBM, por exemplo, apresentou taxas de erro substancialmente mais altas para mulheres negras em comparação com homens brancos, problema identificado como resultado da sub-representação de imagens de mulheres negras nos dados de treinamento.

A falha do sistema em reconhecer corretamente esses rostos não só expôs as limitações técnicas, mas também as implicações sociais de implementar tecnologias enviesadas em contextos como segurança e vigilância. Esse desequilíbrio não apenas levanta preocupações sobre privacidade e vigilância, mas também sobre como essas tecnologias podem reforçar discriminações sistêmicas quando voltadas para contextos de aplicação da lei. Em uma análise recente conduzida por cientistas do Instituto de Tecnologia de Massachusetts (MIT), foi revelado que rostos de pessoas negras são sub-representados nos bancos de dados utilizados para “treinar” esses sistemas. O estudo também mostrou que, em 34,7% dos casos, os algoritmos erroneamente identificaram mulheres negras como homens, enquanto a taxa de erro para homens brancos foi de menos de 1%.

As incertezas sobre a aplicação da tecnologia estão motivando diversas companhias a se distanciarem dela, paralisando um setor que tinha o potencial de alcançar uma movimentação de 7 bilhões de dólares até 2024. A IBM liderou essa mudança ao anunciar, no começo de junho desse ano, sua decisão de cessar o desenvolvimento de sistemas de vigilância em larga escala que comprometam os direitos e as liberdades civis. Em seguida, a Amazon, sob a liderança de Jeff Bezos, suspendeu o uso de seu *software* de reconhecimento facial, Rekognition, pelas forças policiais por um período de um ano. No dia 11 de junho de 2023, a Microsoft também adotou uma postura semelhante, restringindo o acesso a suas ferramentas tecnológicas e declarando que não comercializaria tal tecnologia até que uma legislação federal específica fosse estabelecida.

O *software* da Amazon estava sob escrutínio desde 2023. Um estudo pioneiro realizado pela União Americana pelas Liberdades Civis avaliou a precisão do algoritmo ao comparar fotos dos 535 membros do Congresso com imagens de 25.000 criminosos registrados em um banco de dados. O resultado foi notavelmente errôneo: 28 legisladores foram falsamente identificados como criminosos, apesar de nenhum deles estar em fuga da lei.

Estudos adicionais descobriram que máscaras faciais, como as utilizadas para prevenir a propagação do Covid-19, conseguem enganar os sensores, e há suspeitas de que lentes de contato possam ter efeitos similares. Esse declínio na confiança na tecnologia de reconhecimento facial sublinha como a confiança incondicional frequentemente depositada em inovações – presumidas como baseadas em princípios técnicos rigorosos – pode, em várias ocasiões, representar um significativo mal-entendido.

A Amazon descontinuou o uso de sua ferramenta de recrutamento ao descobrir que o programa desfavorecia candidaturas mencionando a palavra “mulheres” e excluía graduados de instituições femininas. Além disso, critérios como proximidade da residência ao local de trabalho podem involuntariamente discriminar candidatos de regiões mais afastadas ou economicamente desfavorecidas, reforçando barreiras sociais e econômicas já existentes.

Observa-se que, nesse contexto, o racismo algorítmico como manifestação do enraizamento do racismo em sociedades como a brasileira não é apenas um problema técnico, mas fruto de questões profundamente entranhadas nas estruturas sociais e históricas, quando os algoritmos, ao invés de serem ferramentas de objetividade, carregam em si os preconceitos e as visões de mundo de seus criadores.

Nesse sentido, Silva (2022, p.1350) aponta o equívoco da imparcialidade tecnológica, especialmente no que se refere a sistemas de reconhecimento facial. Desde aplicações cotidianas em smartphones até sistemas complexos usados em contextos prisionais, essas tecnologias são vistas como uma reiteração moderna de um projeto colonial, que correlaciona beleza e inteligência exclusivamente com características eurocêntricas. A concepção de visão computacional é influenciada por critérios estabelecidos a partir de distinções étnico-raciais, como demonstrado pela “erotização da identidade negra” em grandes motores de busca como o Google e na popularidade de aplicativos que promovem a alteração da etnia facial, sugerindo um ideal de embranquecimento.

A aplicação dessas tecnologias promete oferecer soluções inovadoras para lidar com os desafios enfrentados pelas instituições responsáveis pela garantia da segurança dos cidadãos. No entanto, como constatado a partir dos exemplos anteriores, o uso da IA na segurança pública também traz consigo uma série de questões éticas, legais e sociais que precisam ser cuidadosamente examinadas nos foros qualificados.

Um dos principais problemas relacionados ao reconhecimento facial é a possibilidade de promoção de injustiças, pois a precisão dos algoritmos utilizados pode variar dependendo de diversos fatores, como a qualidade das imagens utilizadas e a diversidade dos rostos presentes nos bancos de dados. Isso pode levar a erros de identificação e resultar na criminalização indevida de inocentes. Além disso, há evidências de que certos grupos populacionais, como mulheres e representantes de etnias minoritárias, são mais propensos a serem erroneamente identificados ou sub-representados nos sistemas de reconhecimento facial.

Um estudo recente lançado nos Estados Unidos indicou que americanas, asiáticas e negras têm uma chance 100 vezes maior de serem erroneamente identificadas pelo reconhecimento facial do que um homem branco. Tal estudo desencadeou um movimento pela sociedade e pelo congresso americano para iniciar uma discussão ética e legislativa sobre o uso de reconhecimento facial nos Estados Unidos.

Ademais, a NIST (The National Institute of Standards and Technology), responsável por estabelecer os padrões para as novas tecnologias, testou mais de 180 algoritmos de reconhecimento facial de 99 empresas, incluindo gigantes do mercado como Microsoft, Intel e Panasonic. Esses testes revelaram a imperfeição da tecnologia, gerando vieses preconceituosos. Finalmente, a pesquisadora Joy Boulamwini, do MIT MEDIA LAB, tem se manifestado incisivamente contra o viés discriminatório dos algoritmos de reconhecimento facial, criticando duramente vários sistemas, como o da Amazon, que possui um grau de assertividade significativamente menor para mulheres negras.

Logo, o que se constata é que a ineficiência e a imprecisão associadas ao uso do reconhecimento facial têm implicações profundas e potencialmente danosas. Sob o pretexto de aprimorar a segurança por meio de avanços tecnológicos, há uma propensão à discriminação contra grupos minoritários e à promoção de comportamentos excessivos por parte das forças de segurança. Uma perigosa manifestação dessa problemática reside no denominado viés de confirmação do algoritmo. Esse viés surge quando a inteligência artificial é empregada para determinar áreas que necessitam de maior policiamento. Com a maior vigilância em áreas identificadas pelos dados, novos dados são gerados e reintegrados ao sistema, criando um ciclo retroalimentado de viés. Em 2017, o legislador de Nova Iorque

implementou uma regulação visando a transparência dos algoritmos utilizados pelos departamentos públicos da cidade, incluindo aqueles empregados pela polícia.

Os erros no reconhecimento facial podem, evidentemente, provocar a prisão de pessoas inocentes, sobretudo em países marcados pela cultura do encarceramento, que, segundo Silva (2022a, p.110), é vista como uma resposta aos desviantes, corroborando com a visão de Achille Mbembe de que o racismo é uma tecnologia que regula a distribuição da morte e permite as funções assassinas do Estado. A construção de castas raciais em países como Brasil e Estados Unidos resultou em diferentes relações com a polícia e o encarceramento entre os diversos grupos sociais. Nesses contextos, essas tecnologias são projetadas para manter e promover hierarquias sociais de exploração, perpetuando o racismo estrutural ao associar constantemente a imagem do negro com perigo e criminalidade.

Para corroborar essa percepção, Silva (2022a, p.103) descreve um jovem de 14 anos do Complexo da Maré, Rio de Janeiro, Marcus Vinícius, que foi baleado por policiais e morreu tragicamente. Sua mãe, Bruna da Silva, defendeu a inocência de seu filho destacando que ele estava com uniforme escolar e carregava uma mochila com cadernos. O incidente é apresentado como um exemplo da negação da complexidade simbólica de negros e do genocídio negro no Brasil, sustentado por ferramentas ideológicas de supremacia branca e pela mão do Estado. A dissonância social leva os negros brasileiros a um esforço maior para se adaptarem aos padrões brancos, resultando em uma vida sem referências ou parâmetros sólidos. Como argumentado por Silva (2022a, p.105), na era da tecnologia racializada, os algoritmos de decisão automatizada, sustentados por modelos de *machine learning*, contribuem para ocultar e perpetuar as assimetrias sociais, tornando-as menos visíveis e mais difíceis de contestar.

A origem colonial da necropolítica e a imaginação carcerária são exploradas através de um resgate histórico que destaca a normalização da hipervigilância e do controle violento de grupos específicos.

Em 2020, Danillo Félix de Oliveira, educador, foi injustamente preso sob a acusação de roubo. Após dois meses de encarceramento, ele conseguiu comprovar sua inocência e foi liberado. No entanto, o pesadelo voltou a assombrá-lo. Atualmente, Danillo enfrenta novamente um processo judicial baseado em reconhecimento facial, no Estado do Rio de Janeiro. Mesmo com a própria vítima do assalto admitindo o “erro”, a fotografia de Danillo permaneceu armazenada no banco de dados de

suspeitos da 76ª DP, em Niterói. Agora, ele responde à nova acusação em liberdade, diferente do ocorrido em 2020, quando foi detido em plena via pública e passou dois meses preso, sendo transferido entre três presídios distintos.

Figura 2: Imagem de Danillo Félix de Oliveira usada no método de reconhecimento facial.

Danillo questiona reconhecimento facial, método que o tornou réu pela segunda vez. Foto: Reprodução/TV Globo.

Em face dessa flagrante injustiça, o Instituto de Defesa da População Negra assumiu a responsabilidade pela defesa de Danillo. Ele, por sua vez, questionou o método que o colocou novamente como réu: “A situação é idêntica. Parece que, no mesmo dia, essa pessoa [o verdadeiro assaltante] cometeu outros crimes. Os processos são iguais, a juíza é a mesma, a vara também. Eu já fui julgado. Por que preciso passar por outra audiência?”, declarou o educador à imprensa. Danillo, um jovem negro de origem humilde e morador de uma comunidade, expressou sua indignação: “Não havia nenhuma prova para me prenderem, mesmo assim fui preso. Isso não é racismo? Claro que é. Não de uma pessoa em particular, mas do Estado, deste Brasil”.

De acordo com o Ministério Público do Rio de Janeiro (MPRJ), Danillo foi denunciado com base no boletim de ocorrência registrado na delegacia, e as vítimas seriam ouvidas em audiência. O Tribunal de Justiça confirmou que ele é réu e responde por roubo majorado. Segundo dados da Rede de Observatório da Segurança, 90,5% das prisões realizadas por meio de reconhecimento facial envolvem sujeitos negros. Muitos dos detidos nunca tiveram antecedentes criminais e

sequer sabiam como suas imagens foram parar no banco de dados de suspeitos. Apenas 9,5% das prisões envolvem pessoas brancas.

Em junho de 2022, em Salvador, Bahia, um homem foi preso ao chegar a uma festa junina promovida pela prefeitura, no Parque de Exposições. No momento, o homem estava acompanhado da esposa e do filho e foi detido com base na similaridade, contida no banco de dados de reconhecimento facial, entre ele e a pessoa que deveria ser presa. O homem, que é vigilante e nunca havia cometido um crime, foi preso por roubo e ficou 26 dias encarcerado. A Bahia é uma referência mundial no uso de equipamento de reconhecimento facial na área da segurança pública. A secretaria de segurança pública do Estado, no entanto, nunca explicou como a imagem do vigilante foi parar no banco de dados da ferramenta de inteligência artificial, já que ele não possuía antecedentes criminais.

Um levantamento da Defensoria Pública do Rio de Janeiro revelou que 80% dos réus absolvidos por erros no reconhecimento fotográfico passaram, em média, um ano e dois meses presos antes do julgamento. A pesquisa analisou 242 processos julgados pelo Tribunal de Justiça do Estado entre janeiro e junho de 2021. Em um caso extremo, um acusado ficou quase seis anos preso preventivamente antes de ser absolvido. Os erros no reconhecimento fotográfico foram atribuídos ao uso de álbuns de suspeitos, que incluem imagens de pessoas registradas na delegacia ou obtidas em redes sociais. Na maioria dos casos, os sujeitos presentes nesses álbuns não possuem qualquer vínculo com crimes perante a Justiça.

O estudo também destacou o perfil predominante dos acusados identificados por reconhecimento fotográfico: homens negros. Entre os réus julgados, 95,9% são homens, e 63,74% são negros, considerando pretos e pardos conforme a classificação do IBGE. O legado colonial e o supremacismo branco continuam invisíveis em instituições modernas como prisões e leis, perpetuando uma vigilância e uma classificação social racializada que sustenta o capitalismo global. É possível estabelecer paralelos entre o controle colonial e a hipervigilância moderna, diante da persistência de diferenciação racial em documentos de identificação e da suspeição generalizada sobre negros e pobres.

Portanto, a modernização do racismo científico-colonial em áreas como frenologia e criminologia é acompanhada pela continuação da prática biométrica na perseguição de negros. No Brasil, o Estado faz uso das leis e da infraestrutura policial para estratificar a sociedade e promover vigilância constante sobre grupos

específicos. Há uma relação simbiótica entre o aparato policial militar do Brasil do século XIX e as elites escravocratas, que reitera o controle e o genocídio das populações subalternas. Segundo Silva,

> Compreender as tecnologias carcerárias algorítmicas como a distribuição do reconhecimento facial passa por compreender que a "imaginação carcerária" vigente em países moldados pelo colonialismo e pelo supremacismo branco exige entender "*quem e o que* é *fixado no mesmo lugar* – classificado, encurralado e/ou coagido" e como as tecnologias e instituições são criadas para a manutenção e a promoção das hierarquias sociais de exploração (2022a, p.110 – grifo do autor).

Para Benjamin (2019, 2021), ao se considerar raça em si como tecnologia, como um modo de ordenar, organizar e desenhar uma estrutura social assim como a compreensão sobre a durabilidade da raça, sua consistência e adaptabilidade, pode-se entender mais claramente a arquitetura literal de poder. A raça é uma tecnologia designada para estratificar e santificar injustiças sociais, que tem a capacidade de criar universos sociais paralelos e produzir a morte prematura para grupos racializados. As ideologias supremacistas brancas influenciaram a evolução dos meios de comunicação e informação, incorporando uma "imaginação carcerária" na cultura, conforme conceituado por Benjamin (2016).

A autora aborda o conceito de "imaginação carcerária" (2016, p.146) como uma crítica às tecnologias e às políticas que, embora aparentem objetividade, estendem práticas sociais racistas de maneira enganosa. Em sua obra *Captivating Technology*: *Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life*, Benjamin examina como as tecnologias carcerárias são utilizadas para classificar e coagir populações específicas e questiona se essas inovações podem ser resistidas e reimaginadas para fins mais libertadores.

Ela introduz a expressão *The New Jim Code* para explorar uma gama de designs discriminatórios que codificam a desigualdade, argumentando por uma compreensão ampla do "carceral" que vai além da polícia, incluindo formas de contenção que tornam a inovação possível em contextos de saúde, educação, emprego, políticas de fronteira e realidades virtuais. No contexto estadunidense do qual Benjamin faz parte, o racismo institucional opera como uma tecnologia social que influencia profundamente a imaginação carcerária que apoia o sistema penal americano.

A vigilância policial, baseada em avaliações algorítmicas de risco "criminal", utiliza perfis construídos com base em dados criminais, educacionais, de emprego,

econômicos, geográficos e outros aspectos moldados pela dominação racial. A marginalização dos negros, inerente a esse perfilamento algorítmico, é vista como um elemento de risco. Assim, as tecnologias carcerárias de vigilância perpetuam e intensificam a dominação racial sobre os negros (BENJAMIN, 2019).

Benjamin (2021, p.22) mostra como a imaginação carcerária na tecnologia eleva a outro nível o que Fanon denomina de epidermização racial, ou seja, a imposição do peso da raça ao corpo. Com o avanço tecnológico, ela passa a ser uma epidermirzação digital, que, por meio da renderização de códigos racialmente embasada, através do reconhecimento facial, tomografia de íris e retina, geometria da mão, modelos de impressões digitais, padrões vasculares, de marcha e outros reconhecimentos cinéticos e do DNA, além de outras formas de codificação corporal, tende a projetar o corpo negro como uma evidência de um crime que pode vir a ocorrer. A criminalização do corpo negro toma um passo à frente, penalizando a negritude por meio de uma ferramenta tecnológica que reforça o encarceramento em massa.

7.3.2 Testes padronizados usados na saúde e sistema de pontuação de crédito e seleção de emprego

Eubanks (2018) assevera que a automação de serviços sociais nos Estados Unidos, por exemplo, afeta de maneira desproporcional as minorias raciais e as populações de baixa renda, resultando em novas formas de exclusão e precariedade. Isso evidencia que o racismo algorítmico tem implicações diretas nas vidas dos sujeitos, estabelecendo novas formas de governança baseadas em dados e automação.

Nesse contexto, ferramentas de inteligência artificial têm sido utilizadas na saúde para definir a prioridade de atendimento médico. Um dos elementos que alimentam e treinam os sistemas é a raça, e esse critério como fator para decidir sobre cuidados de saúde tem sido frequente, o que pode aumentar o impacto da discriminação racial nos cuidados de saúde, especialmente ao considerar que esses aplicativos aprendem por meio de dados para funcionar automaticamente. Um estudo de 2019 demonstrou que pacientes negros considerados com quadros graves são elencados no mesmo nível de risco que pacientes com quadros medianos em um aplicativo que serve para a gestão de cuidados de alto risco. O algoritmo em questão,

utilizado comercialmente na área da saúde dos Estados Unidos, utiliza como critério a correlação entre os custos dos cuidados de saúde e a raça, ao invés da gravidade da doença, *ipsi litteris*, para definir o risco.

Os algoritmos são cada vez mais utilizados para diagnosticar doenças, prever riscos de saúde e personalizar tratamentos. No entanto, quando esses modelos são treinados predominantemente com dados de populações específicas, a sua aplicabilidade e a sua precisão para outras populações podem ser significativamente comprometidas. Isso pode levar a diagnósticos incorretos ou tratamentos inadequados para grupos sub-representados, exacerbando disparidades em saúde. Algoritmos utilizados para triagem e tratamento de pacientes (SILVA, 2020; EUBANKS, 2018; CARNEIRO, 2005; BENJAMIN, 2016) têm mostrado viés ao subestimar a severidade das condições de pacientes não brancos, impactando negativamente o acesso a tratamentos e recursos médicos necessários. Esse problema é exacerbado pela falta de diversidade nos conjuntos de dados de treinamento, que frequentemente não representam de maneira adequada populações racialmente diversas, levando a decisões clínicas que podem ter consequências fatais.

Nos Estados Unidos, o sistema de saúde adota algoritmos comerciais para auxiliar nas decisões relacionadas aos cuidados médicos. Obermeyer e colaboradores identificaram a presença de preconceito racial em um desses algoritmos amplamente utilizados. Foi constatado que pacientes negros, classificados no mesmo nível de risco que pacientes brancos, apresentavam condições de saúde mais graves. Segundo os pesquisadores, esse viés reduz pela metade o número de pacientes negros selecionados para receber cuidados adicionais. Os sistemas de saúde utilizam algoritmos comerciais de previsão para identificar e oferecer suporte a pacientes com condições de saúde complexas.

Contudo, evidencia-se que um algoritmo amplamente empregado, comum em todo o setor e impactando milhões de pessoas, apresenta um viés racial significativo. Dentro de uma mesma classificação de risco, pacientes negros frequentemente apresentam condições de saúde mais graves do que pacientes brancos, conforme demonstrado por sinais de doenças não controladas. A correção dessa desigualdade poderia aumentar a proporção de pacientes negros beneficiados por cuidados adicionais de 17,7% para 46,5%.

Esse viés ocorre porque o algoritmo considera os custos de assistência médica como indicador das necessidades de saúde. No entanto, o acesso desigual aos cuidados faz com que menos recursos sejam destinados a pacientes negros, ainda que tenham necessidades semelhantes às de pacientes brancos. Assim, embora os custos de assistência médica possam parecer uma medida útil para prever a saúde em alguns casos, eles geram grandes distorções raciais.

O problema ocorre porque o algoritmo utiliza os custos de saúde como indicador substituto das necessidades de cuidado. Como menos recursos financeiros são destinados aos pacientes negros, mesmo quando possuem as mesmas necessidades que os brancos, o algoritmo interpreta equivocadamente que os pacientes negros estão em melhores condições de saúde. Ao ajustar o algoritmo para que ele não utilize os custos como referência das necessidades de saúde, é possível eliminar o preconceito racial na identificação de quem necessita de cuidados extras.

Além do campo da saúde, a integração de algoritmos também se faz presente nos processos decisórios relativos ao acesso a serviços essenciais e crédito, conforme Da Rocha *et al.* (2020, p.6), representa uma tendência crescente na era digital. Embora essa integração prometa maior eficiência e objetividade, ela também suscita preocupações significativas sobre a perpetuação de desigualdades sociais e econômicas, particularmente no que diz respeito à discriminação racial.

Esse fenômeno manifesta-se de maneira preocupante nos domínios da habitação e do crédito, em que as decisões algorítmicas podem ter impactos profundos no corpo social, especialmente em comunidades racialmente minoritárias. Nesses sistemas, algoritmos são frequentemente utilizados para avaliar a solvência e determinar a elegibilidade para empréstimos, hipotecas e outras formas de crédito. No entanto, esses sistemas podem incorporar vieses que resultam em avaliações desfavoráveis para sujeitos pertencentes a minorias raciais.

Isso ocorre, em parte, devido a conjuntos de dados históricos que refletem desigualdades sistêmicas, levando os algoritmos a perpetuarem padrões de exclusão ao invés de neutralizá-los. Na habitação, o uso de algoritmos para tomar decisões sobre locação e venda de propriedades pode também reforçar barreiras existentes. A automação dos processos de triagem de inquilinos e compradores, baseada em critérios que podem indiretamente refletir preconceitos raciais, amplia o risco de discriminação.

Silva (2020, pp.129-130) afirma que os sistemas de pontuação de crédito oferecem outro exemplo notável, no qual o racismo algorítmico se manifesta através da inclusão de variáveis que, embora aparentemente neutras, refletem desigualdades socioeconômicas raciais. Fatores como localização geográfica e histórico de emprego podem indiretamente codificar informações raciais, resultando em pontuações de crédito mais baixas para grupos pertencentes a minorias étnicas. Essas pontuações, por sua vez, afetam a capacidade desses sujeitos de obter empréstimos, hipotecas e outros serviços financeiros, perpetuando ciclos de desvantagem econômica. Algoritmos têm determinado a elegibilidade para empréstimos com base em dados históricos que refletem desigualdades raciais profundas, resultando em taxas de aprovação mais baixas para candidatos de grupos raciais minoritários, perpetuando, desse modo, ciclos de pobreza e exclusão econômica (DA SILVA, 2020; VILELA, 2011; DA ROCHA *et al.*, 2020).

Esse uso de algoritmos na decisão de crédito ilustra como a suposta neutralidade dos sistemas automatizados pode, na verdade, fortalecer barreiras sociais e econômicas existentes. Algoritmos utilizados por instituições financeiras para avaliar a solvência de pessoas, conforme Oliveira (2023), muitas vezes desfavorecem minorias raciais e comunidades de baixa renda. Essas disparidades decorrem do uso de variáveis *proxy*[22], como código postal, que estão correlacionadas com *status* socioeconômico e etnia. Como resultado, grupos dessas comunidades enfrentam barreiras adicionais ao acesso a empréstimos e serviços financeiros, bem como a empregos.

Nesse sentido, a cada dia mais influenciado por novas tecnologias digitais, o mercado de trabalho tem integrado sistemas algorítmicos no processo de recrutamento e seleção de candidatos para vagas de emprego. Nesse contexto, a ferramenta virtual promete ser imparcial e mais eficiente. No entanto, também desperta o alerta de que ela possa perpetuar desigualdade racial, quando impede a inclusão de minorias no mercado de trabalho. Algoritmos para filtrar candidatos e realizar análise de currículos foram automatizados, o que em tese poderia eliminar o viés humana desses processos. Porém, mais uma vez, os dados e os mapas utilizados para programar sistemas de recrutamento podem reproduzir e até mesmo amplificar o viés existente em relação à raça.

[22] Uma variável *proxy* é aquela que se apresenta no lugar da real variável de interesse, a qual não pode estar disponível, ser muito cara ou muito demorada de medir (BRUC & BRUCE, 2019).

Quando o sistema é treinado com dados históricos que refletem um desequilíbrio de gênero ou raça, podem reproduzir e exacerbar desigualdades. Por exemplo, algoritmos que penalizam palavras ou experiências associadas a grupos sub-representados podem impedir que candidatos qualificados dessas populações obtenham oportunidades de emprego[23], sendo que sistemas de triagem de currículos (SOUSA, 2022; LEE, 2019; SWEENEY, 2013) programados para identificar candidatos “ideais” podem filtrar nomes que soem “não tradicionais”, um eufemismo frequentemente usado para descrever nomes comuns em comunidades de minorias raciais. Por exemplo, um algoritmo de análise de currículos pode aprender a associar certos nomes com qualificações inferiores, não porque os desenvolvedores programaram explicitamente essa associação, mas porque os dados históricos refletem uma discriminação sistemática.

A ferramenta, sistema de contratação automatizada, desenvolvida pela Amazon para triagem de currículos apresentou vieses de gênero. O sistema, treinado com currículos enviados à empresa ao longo de dez anos, começou a penalizar candidaturas que continham termos associados a mulheres, como *women’s* (referente a organizações de mulheres). Esse viés surgiu porque a maioria dos currículos anteriores vinha de homens, refletindo a composição historicamente desequilibrada da força de trabalho na área de tecnologia. Tal prática ilustra como o racismo algorítmico pode limitar oportunidades de emprego para essas comunidades, ampliando desigualdades pré-existentes no mercado de trabalho.

Esses algoritmos são treinados em dados historicamente constituídos que, frequentemente, incluem discriminação contra certos perfis que já não são amplamente privilegiados em suas respectivas áreas de atuação. Os sistemas de triagem, por exemplo, podem excluir automaticamente candidatos de grupos sub-representados, em áreas dominadas por pessoas relacionadas a grupos majoritários, o que perpetuará o ciclo de exclusão e limitará ainda mais o acesso das minorias ao mercado de trabalho.

Nos processos seletivos os quais utilizam inteligência artificial para analisar entrevistas, por exemplo, as avaliações podem ser baseadas em aspectos como padrão de fala, expressões faciais e linguagem corporal, que nem sempre são culturalmente neutros, culminando na exclusão de candidatos de minorias raciais,

[23] El algoritmo de Amazon al que no le gustan las mujeres. Disponível em: https://www.bbc.com/mundo/noticias- 45823470. Acessado em: 12/08/2024.

cujas formas de comunicação podem diferir dos padrões considerados “ideais” pelos algoritmos. Em decorrência disso, surgem obstáculos adicionais, os quais impedem o acesso desses grupos ao mercado de trabalho. O racismo algorítmico, enquanto prática recorrente, por atuar de maneira discreta, torna-se ainda mais preocupante, uma vez que os candidatos estão cada vez mais alheios aos processos que influenciam nas decisões de contratação.

Nos processos de seleção automatizados utilizados em contextos de recrutamento e admissões acadêmicas, algoritmos podem favorecer candidatos que se encaixam em perfis históricos de sucesso, muitas vezes baseados em dados que refletem vieses raciais e socioeconômicos. Por exemplo, algoritmos podem valorizar características ou experiências que são mais acessíveis a candidatos de *backgrounds* mais privilegiados, excluindo, assim, talentos potenciais de comunidades marginalizadas (SILVA, 2020, p.128). Essa prática não apenas limita as oportunidades para aqueles pertencentes a esses grupos, mas também reforça estruturas institucionais de desigualdade, demonstrando como os algoritmos podem ser instrumentos de exclusão social.

7.3.3 Sistemas de branqueamento em filtros de aplicativos; imagens geradas por inteligência artificial (IA) e sistemas de recomendação usados nas redes sociais

O viés algorítmico presente nos atuais sistemas tecnológicos, notadamente digitais, favorece uma eficiência alinhada aos princípios hegemônicos, seguindo a tradição do pensamento colonial. Desse modo, as formas de produção e disseminação das tecnologias digitais contemporâneas podem ser analisadas pelo prisma da colonialidade, que se baseia na exploração contínua de determinados corpos, sujeitos e territórios em benefício de outros, frequentemente sob pretexto de racionalidade, civilização e progresso. Assim, a posição subjugada do corpo negro é prolongada, servindo como um vetor para a perpetuação de estereótipos que sustentam a supremacia branca (CARRERA, 2021, p.11).

Sob a concepção de que o poder não se fixa em um único ponto, mas circula através de diferentes esferas, torna-se possível compreender os mecanismos que sustentam essa lógica, por meio de uma prática que garante a valorização da branquitude como identidade dominante, molda subjetividades e estabelece padrões de exclusão. A colonialidade embutida nos sistemas algorítmicos enviesados revela a

existência de sistemas de poder que formam mercados de trabalho, alteram dinâmicas geopolíticas e constroem o discurso ético, exercendo dominação, não por meio da força física, mas através de mecanismos discretos de controle sobre o ecossistema digital e sua infraestrutura (CARRERA, 2021, pp.8-9).

Nessa mesma perspectiva, os discursos sobre beleza e estética, em linhas gerais, são apresentados como verdades irrefutáveis, criando barreiras e impondo padrões seletivos os quais excluem com base na cor da pele. Esses critérios, usados para definir normas e sustentar discursos preconceituosos, não apenas classificam como também rotulam sujeitos, como estruturas de poder na comunicação que reforçam, de modo sutil ou não, a discriminação racial por meio de filtros de aplicativos, sistemas de recomendação e imagens geradas por IA, os quais revelam como padrões estéticos podem ser utilizados como um instrumento de exclusão e reforço de desigualdades sociais.

Isso se dá porque os discursos nunca são destituídos de efeitos de sentido e efeitos de verdade, as representações imagéticas constroem modelos, que favorecem o preestabelecimento de padrões, notadamente aqueles relativos a questões estéticas sobre cabelo, tom da pele e traços faciais, por exemplo. Nesse processo, o sujeito absorve e reproduz discursos digitais que naturalizam perspectivas baseadas na branquitude, transformando o branqueamento em paradigma incontestável através de mecanismos discursivos historicamente construídos.

Esta investigação analisa como os vieses presentes nos algoritmos, fundamentados em estruturas de poder e práticas sociais constituintes de identidades, materializam o que Foucault (2013) denominou como o movimento discursivo de retomada do já conhecido para produzir novas significações. A doutrina da estética do corpo em si se constituiu como discurso recorrente notadamente na pós-modernidade primeiramente pelo Photoshop, depois em forma de programas de edição de imagem como Adobe Photoshop, Lightroom e aplicativos de retoque facial como Facetune e Snapseed, usados para alterar tons de pele ou reforçar padrões estéticos discriminatórios como o embranquecimento de pessoas negras ou a valorização de traços eurocêntricos proposta pelos discursos da beleza e estética, uma biopolítica contemporânea, como forma de controle e de exclusão.

Nesse sentido, é importante atentar para as seguintes considerações de Foucault:

> À primeira vista, as “doutrinas” (religiosas, políticas, filosóficas) constituem o inverso de uma “sociedade de discurso”: [...] A doutrina, [...], tende a difundir-se; e é pela partilha de um só e mesmo conjunto de discursos que indivíduos, tão numerosos quanto se queira imaginar, definem sua pertença recíproca. Aparentemente, a única condição requerida é o reconhecimento das mesmas verdades e a aceitação de certa regra – mais ou menos flexível – de conformidade com os discursos validados; [...] a doutrina questiona os enunciados a partir dos sujeitos que falam, na medida em que a doutrina vale sempre como o sinal, a manifestação e o instrumento de uma pertença prévia – pertença de classe *status* social ou de raça, de nacionalidade ou de interesse, de luta, de revolta, de resistência ou de aceitação. A doutrina liga os indivíduos a certos tipos de enunciação e lhes proíbe, consequentemente, todos os outros, mas ela se serve, em contrapartida, de certos tipos de enunciação para ligar indivíduos entre si e diferenciá-los, por isso mesmo, de todos os outros (2013, pp.39-40).

Isso implica perceber que há sim uma espécie de doutrinação no sentido de conduzir os sujeitos a aceitarem o discurso dominante, algo refletido em aplicativos como o FaceApp, que nitidamente produz um processo de branqueamento de sujeitos negros, conforme se vê no registro a seguir:

Figura 3: *Print* do aplicativo FaceApp que “branqueia” os usuários para torná-los “mais sexy”

O efeito do filtro do *FaceApp* para se ver “mais sexy”. *Twitter* (El País. Madri-25 ABR 2017).

O aplicativo FaceApp, cuja configuração possibilita que o usuário altere fotos do rosto, tornando a imagem envelhecida, branqueou a pele de um usuário, quando este configurou o App para ficar mais sexy. Os criadores do aplicativo, após repercussão negativa, justificaram se tratar de “um efeito secundário infeliz” da tecnologia utilizada nos filtros. O fato descrito aconteceu com inúmeros usuários negros. Shahquelle L., um rapaz de 21 anos, relatou que a ferramenta não só é ruim, é também racista. A ideia de beleza foi associada a pele clara, cabelos lisos, olhos claros e nariz afilado, evidenciando que o algoritmo reproduziu um padrão normativo histórico do que “socialmente” se perpetua, equiparando beleza à brancura, seguindo um padrão eurocêntrico, em oposição a traços afrodescendentes, por exemplo.

O enviesamento algorítmico não se dá isoladamente, pois podem-se enumerar casos descritos, em *sites*, revistas eletrônicas e artigos diversos, em que a inteligência artificial compreende e interpreta imagens de negros portando alguns objetos (por exemplo, termômetro, furadeira, pincel) como marginais segurando armas.

Os alcançados a partir do FaceApp revelam que os mecanismos discursivos do branqueamento permanecem ativos no imaginário social contemporâneo. Esses discursos, que se mantêm persistentes na esfera pública, continuam a ressoar na memória coletiva, reativando e reforçando padrões estabelecidos. Os efeitos de sentido desencadeados são desoladores, uma vez que essa associação de beleza à branquitude inscreve-se numa discursividade discreta, um construto histórico-ideológica de memória, ativado por mecanismos que disseminam a ideia de belo a partir da branquitude.

Ao examinar como o branqueamento se manifesta nas interações diárias, particularmente no ambiente digital, através dos atuais padrões de beleza, podem-se observar seus impactos na construção da memória coletiva. Esse fenômeno ganha força ao se apresentar como algo único e, portanto, válido por natureza. Segundo Hofbauer (2003), a linguagem discursiva, em sua natureza não-transparente, serve como veículo para práticas de branqueamento que se camuflam sob a autoridade de discursos pretensamente científicos, o que provoca efeitos de sentido de exclusão do negro nessa arena da estética padronizada, notadamente, europeizada e que obviamente não corresponde a traços e feições dos povos africanos ou racializados.

O *whitening phenomenon,* que é a utilização de filtros em aplicativos de redes sociais para clarear a pele, tem se configurado como uma manifestação contemporânea do racismo estrutural. Essa prática, além de ser bem pouco discutida com seriedade na literatura, apresenta ramificações maiores do que usualmente consideradas quanto às identidades dos usuários, uma vez que promove uma idealização da pele clara e de características faciais associadas à branquitude, explicitando modelos através dos quais as novas tecnologias digitais favorecem as normatizações e as normalizações de discursos enviesados.

A origem desse fenômeno pode ser traçada até as raízes históricas do colonialismo e do racismo científico, que estabeleceram a supremacia branca como um ideal universal. Essa prática agora encontra um novo meio de expressão na era digital, através de filtros que clareiam a pele, afinam o nariz e alteram os traços de forma a alinhá-los com padrões eurocêntricos. Não são apenas ferramentas de edição

de imagem, mas também instrumentos de normalização de uma estética racial específica que reflete uma internalização de padrões de beleza que valorizam características associadas à branquitude, ao mesmo tempo em que marginalizam e estigmatizam características fenotípicas de outros grupos raciais.

Essa dinâmica contribui para a construção de uma hierarquia de beleza que posiciona a branquitude no ápice, incentivando processos de desracialização entre sujeitos não-brancos em busca de aceitação social e validação dentro de plataformas digitais. Nas palavras de Foucault,

> a raça, o racismo é a condição de aceitabilidade de tirar a vida numa sociedade de normalização [...]. A função assassina do Estado só pode ser assegurada, desde que o Estado funcione no modo do biopoder, pelo racismo [...]. É claro, por tirar a vida não entendo simplesmente o assassínio direto, mas também tudo o que pode ser assassínio indireto: o fato de expor à morte, de multiplicar para alguns o risco de morte ou, simplesmente, a morte política, a expulsão, a rejeição, etc. (2021b, pp.215-216).

Essa prática racista se pereniza na atualidade pela reprodução de um discurso que legitima o branqueamento enquanto norma, articulando estratégias de poder que permeiam o cotidiano – desde os ideais de beleza até as estruturas estéticas –, reproduzindo desigualdades no corpo social.

Operando como estratégia discursiva legitimada por autoridades, respaldada pela ciência e funcionando como um mecanismo de poder bem estruturado, os repositórios de imagens funcionam como manifestação de estruturas de poder de cunho colonial, evidenciadas tanto pela imprecisão na representação de determinados grupos sociais quanto pelo apagamento em certos contextos significativos e em sua exposição excessiva em outros.

Assim, os estereótipos raciais atuam como ferramentas simbólicas para perpetuar o domínio da supremacia branca e a subalternidade imposta aos negros, sublinhando a ideia de que o poder soberano se baseia na demarcação da ameaça que o “outro” representa. Esses sistemas digitais operam por meio de processos detalhados de etiquetagem, seleção e definição de relevância, gerenciando como as palavras-chave são vinculadas às imagens em seu acervo e como certos resultados são priorizados em detrimento de outros, baseando-se em critérios de popularidade ou conexões semânticas (CARRERA, 2021, p.29).

A falta de transparência sobre como seus algoritmos operam contribui para uma efetiva obscuridade em sua base tecnológica, pois incorporam valores implícitos nas decisões automatizadas sem explicitar as dinâmicas que moldam tais escolhas.

Plataformas como Shutterstock e Getty Images são repositórios que utilizam métodos algorítmicos enviesados de modo a determinar a relevância de certos conteúdos, gerando impactos nas escolhas de reprodução de padrões coloniais na era atual.

Estudos anteriores, segundo Carrera (2021, p.11), já demonstraram como esses repositórios podem corroborar na perpetuação de desvalorizações afetivas direcionadas a mulheres negras e intensificar estereótipos ao associar atributos negativos ao corpo negro e positivos, ao corpo branco, revelando a complexidade e a opacidade algorítmica que oculta seus critérios de relevância.

Benjamin (2019, p.145) destaca que a amplificação dos processos discriminatórios presentes na sociedade e replicados por algoritmos e sistemas de inteligência artificial está diretamente relacionada à percepção equivocada de que essas tecnologias são neutras e objetivas. Como ela aponta, “a neutralidade algorítmica reproduz discriminação sustentada algoritmicamente” (2019, p.145), evidenciando que essas ferramentas podem reforçar desigualdades em vez de eliminá-las. Silva (2019a) complementa essa análise ao afirmar que essa suposta neutralidade tecnológica gera uma opacidade dupla no funcionamento dos sistemas automatizados. O conceito de dupla opacidade, discutido no âmbito da Teoria Racial Crítica[24] (TRC), para o autor, refere-se à dificuldade de abordar simultaneamente as questões raciais e a forma como valores e preconceitos são incorporados em tecnologias e dispositivos digitais. Segundo a TRC, existem dois níveis de invisibilidade que dificultam uma análise crítica sobre como esses fatores se entrelaçam nos sistemas tecnológicos, perpetuando um discurso enviesado e profundamente enraizado nas estruturas sociais.

O primeiro nível de opacidade refere-se à maneira como discursos dominantes amplamente aceitos, frequentemente, minimizam ou ignoram a importância das questões raciais no campo tecnológico, resultando em uma representação limitada das vivências e das contribuições de grupos racialmente marginalizados na

[24] Criada por juristas e intelectuais negros, em meados da década de 1970, nos Estados Unidos, a Teoria Racial Crítica (*Critical Race Theory* – CRT), a priori, constituiu-se dentro da esfera jurídica, em resposta à incapacidade jurídica no enfrentamento ao racismo sistêmico. Derrick Bell, Kimberlé Crenshaw, Cheryl Harris, Richard Delgado e Mari Matsuda figuram entre os seus principais teóricos, que buscavam demonstrar que o racismo não é apenas um conjunto de preconceitos individuais, mas uma estrutura institucionalizada que molda políticas, relações sociais e econômicas. A TRC propõe uma abordagem crítica ao direito, ao afirmar que o racismo é estrutural, sistemático e profundamente enraizado nas instituições sociais, incluindo o sistema jurídico, e não apenas o resultado de atitudes individuais preconceituosas (DELGADO & STEFANCIC, 2017). Para a TRC, o racismo é parte constitutiva da ordem social, mantendo privilégios para grupos racialmente dominantes.

concepção e no desenvolvimento dessas tecnologias. Já o segundo nível refere-se à dificuldade de reconhecer que as tecnologias não são neutras e podem reproduzir desigualdades sociais, incluindo as raciais, quando sistemas digitais ignoram a diversidade racial e cultural, perpetuando preconceitos e reforçando estereótipos.

Nessa perspectiva, o racismo é tido como elemento estruturante do funcionamento da sociedade, manifestado no dia a dia dos grupos racializados. A raça, por sua vez, não é uma realidade biológica ou genética, mas um construto social, criado, manipulado e descartado conforme interesses e conveniências de determinados grupos. Mulheres negras, em particular, encontram nesses espaços a cristalização de suas marginalizações através de imagens, filtros e aplicativos que reiteram as limitações impostas sobre sua subjetividade pelo discurso colonial inerente às tecnologias digitais.

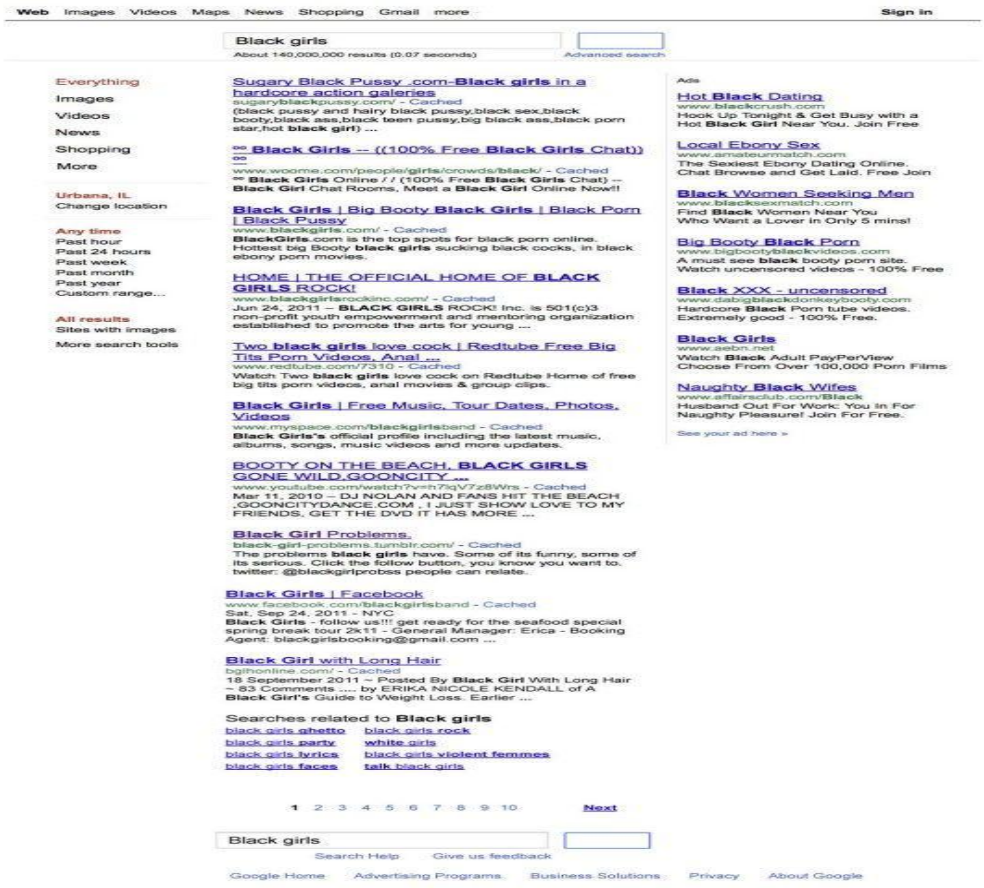
A consequência dessa prática vai além da esfera digital, afetando a percepção que o sujeito tem de si e dos outros no mundo real, contribuindo para a perpetuação de estereótipos raciais, influenciando as expectativas sociais em relação à aparência e reforçando a marginalização de características não eurocêntricas (FERNANDES, 2023).

Sistemas de recomendação usados em plataformas de mídia social, como Facebook e YouTube, imagens gradas por IA e diferentes tipos de aplicativos são espaços em que se pode observar a ocorrência de racismo algorítmico, quando priorizam conteúdo que maximizam discursos enviesados, os quais reiteram a ideia de branquitude, quando estigmatizam a representação de grupos racializados.

Assim, é reconhecido que determinados grupos são afetadas de maneira negativa por decisões automatizadas e treinamentos de algoritmos tendenciosos, evidenciando-se, por exemplo, como certos mecanismos de busca podem ligar mulheres negras a conteúdos indevidos. Portanto, ferramentas automatizadas podem perpetuar desigualdades sociais, e tecnologias de reconhecimento facial podem intensificar a exclusão baseada em gênero e raça (CARRERA, 2021, p.7).

Dentro dessa lógica de opressão interseccional, as mulheres negras são particularmente visadas, sendo racializadas e sexualizadas através da perspectiva do homem branco, que controla as diretrizes de produção desses bancos de imagens e influencia o panorama midiático e comunicacional em larga escala. Isso pode ser conferido na Figura 5, que expõe os resultados de busca nada favoráveis para “garotas negras”:

Figura 4: *Print* do Resultado da busca por “garotas negras”, no Google.



Resultado da pesquisa pela palavra-chave “garotas negras’’, em 18/11/2011. Fonte:www.invisibleculturejournal.com/pub/google-search-ypervisibility#n17i7sa4qef

Esses repositórios de imagens, como ferramentas racistas e sexistas, evidenciam a parcialidade dessas tecnologias e seu papel na perpetuação algorítmica de desigualdades arraigadas em contextos socioculturais diversos, em uma espécie de colonialidade tecnológica.

Renata Souza, deputada estadual pelo PSOL-RJ e líder da CPI do Reconhecimento Fotográfico nas Delegacias na ALERJ, trouxe à tona preocupações sobre o racismo presente em sistemas de inteligência artificial. Ela denuncia ter sofrido racismo algorítmico, em outubro de 2023, após gerar uma imagem por meio de uma inteligência artificial, ao desenvolver uma arte baseada em pôsteres de desenhos animados. Na ocasião, ela emitiu o comando para que a IA fizesse um desenho de uma mulher em uma favela, com base em suas características pessoais. A parlamentar forneceu um *prompt* de comando ao aplicativo para que fosse gerada uma imagem, num cenário de favela, de uma mulher negra, com cabelo afro, trajando indumentária africana, ao estilo dos pôsteres da Disney. Na imagem gerada, a mulher

tinha uma arma de fogo em suas mãos e o cenário da favela ao fundo. Na ocasião, a deputada questionou a relação discursiva, implicada no resultado gerado pelo *software*, entre a negritude e a criminalidade.

Figura 5: *Print* da imagem gerada por IA na *trend* da Disney Pixar.

Imagem gerada por inteligência artificial a pedido de Renata Souza/Foto: Reprodução/X.

De acordo com a parlamentar, em nenhum momento, na elaboração do *prompt* de comando, ela citou armas ou violência. Era para ser gerado um texto icônico sobre uma mulher negra na favela, e, ainda assim, a imagem gerada a representou segurando uma arma. Isso evidencia a presença do racismo algorítmico. A criminalização de negros que vivem em favelas e periferias não ocorre apenas na sociedade, mas também nos sistemas automatizados. Como presidente da CPI do Reconhecimento Fotográfico, ela pôde observar o impacto desses algoritmos nas tecnologias de inteligência artificial e reconhecimento facial, que frequentemente associam negros, pobres e jovens à criminalidade, reforçando um viés discriminatório preocupante.

Os fundamentos dos estudos tomados para análise aqui residem na ideia de que a neutralidade algorítmica não existe, é apenas uma falácia evidenciada na obscuridade das tecnologias e nos preconceitos embutidos na inteligência artificial. Nas questões raciais, observa-se o surgimento do que se pode chamar de “código Jim” contemporâneo, referindo-se ao uso de novas tecnologias para replicar velhas dinâmicas de exclusão racial, promovidas e percebidas como mais objetivas ou avançadas do que os sistemas discriminatórios do passado.

Entende-se a presença de uma “dupla obscuridade” na construção desses dispositivos, pois a discriminação racial resulta tanto da falta de compreensão sobre

a profundidade das desigualdades raciais, embutidas nas práticas e nos comportamentos de programação, quanto da ausência de medidas éticas integradas nos processos de desenvolvimento dessas ferramentas (CARRERA, 2021, pp.7-8). Dessa forma, algumas dessas práticas, por serem sutis e ocultarem seu impacto prejudicial, realizam “microagressões codificadas”, ou seja, formas veladas de perpetuar ofensas raciais tão severas quanto o racismo explícito.

Relativamente à representação imagética, ela está entalhada em uma discursividade constituída por práticas culturais, históricas e ideológicas. Nesse processo, mobilizam-se recursos semióticos para sustentar determinadas representações, que emergem de uma memória discursiva construída historicamente. Essas representações são produzidas por práticas sociais recorrentes, que não apenas descrevem, mas ativamente constituem os próprios objetos que nomeiam – o que caracteriza o funcionamento do discurso em sua dimensão constitutiva. Esses recursos semióticos colaboram para consolidar o padrão da brancura e para a normalização dos sujeitos dentro desse sistema. Esse fenômeno é alimentado por estruturas ideológicas racistas baseadas em hierarquias de cor, que perpetuam relações desiguais de poder em um jogo discursivo em que os sujeitos alternam posições, conforme as estratégias de subjetivação e objetivação que regulam a dinâmica do poder.

Segundo Noble (2021), as plataformas tecnológicas e os sistemas de busca desempenham um papel central na criação de uma “realidade algorítmica” que perpetua estereótipos raciais e de gênero e reforçam hierarquias raciais ao controlar o acesso à informação e definir como diferentes grupos raciais são representados. No exemplo da Figura 4, ao buscar pela palavra-chave “garotas negras”, a autora obteve como resultado de primeira página, em sua maioria, incluindo os cinco primeiros *links*, páginas pornográficas, com títulos e descrições sexualmente explícitos.

Esses resultados refletem uma visão distorcida e sexualizada dessas mulheres negras, criando uma forma de racismo mediado por algoritmos. Esse tipo de prática tecnológica atua como uma máquina produtora de discurso, que valida preconceitos raciais sob o disfarce de neutralidade e objetividade tecnológica. Para a pesquisadora, os algoritmos carregam preconceitos que derivam das estruturas sociais nas quais foram desenvolvidos. Desse modo, a tecnologia se torna um instrumento de poder que contribui para a legitimação e a perpetuação de um discurso racial excludente.

Em bancos de imagens, essa estrutura de poder pode se manifestar na falta de precisão sobre a representação de determinados grupos sociais, no apagamento de alguns sujeitos e na recorrência de outros em determinados contextos de pesquisa através de palavra-chave e uma recorrente estereotipia por meio de discursos discriminatórios e hiperritualizações de comportamentos culturalmente determinados nesses ambientes.

A maioria dos engajamentos com meios digitais nos mecanismos de busca e a quase universalidade das interações nesse cenário se dão por intermédio do Google, adotado e incorporado em variados espaços da rede de modo quase onipresente. No entanto, ao longo do processo de popularização do uso dessa ferramenta, os usuários se depararam com questões críticas relacionadas a valores atribuídos, principalmente à raça e ao gênero em sistemas de classificação e indexação da *web* e os resultados de busca que eles retornam.

Essa *big tech* tem a capacidade de priorizar resultados de pesquisa on-line a partir de uma variedade de interesses. As representações de mulheres (particularmente mulheres negras que são codificadas como “garotas”) são frequentemente classificadas em uma página de mecanismo de pesquisa de maneiras que ressaltam aspetos negativos, depreciativos e misóginos na sociedade. Os resultados da Pesquisa Google sobre as palavras "garotas negras" refletem discursivamente o poder social hegemônico e o preconceito racista e sexista, apontando para um tipo de hegemonia cultural dentro dos resultados sobre identidades racializadas e de gênero. São os discursos dominantes sobre a objetividade e a popularidade dos resultados de pesquisa na *web* que fazem os resultados de pesquisa misóginos ou racistas parecerem naturais.

Além disso, a crença generalizada nos mitos da democracia digital[25] emblemáticos do Google e dos seus resultados de pesquisa normaliza a ideia de que

[25] A democracia digital, baseada na ideia de que a internet ampliaria o acesso à informação e garantiria voz a todos os cidadãos, revela-se uma falácia. Na prática, as plataformas digitais continuam a reproduzir desigualdades sociais, políticas e econômicas, beneficiando determinados grupos enquanto marginalizam outros. A promessa de livre acesso ao conhecimento e à expressão, nos meios digitais, ignora as barreiras socioeconômicas que dificultam a participação cidadã. Segundo Van Dijk (2020) e Noble (2021), fatores como renda, infraestrutura precária e falta de alfabetização digital ampliam formas de exclusão, contrariando o ideal de equidade digital. Para Zuboff (2021) e Silva (2020), os algoritmos dessas plataformas priorizam conteúdos sensacionalistas e desinformação, dificultando o debate plural e crítico. Governos autoritários e corporações utilizam as redes digitais para monitorar, censurar e manipular informações, práticas como o uso de *bots* para influenciar percepções eleitorais e a moderação seletiva de informações evidenciam que o ambiente digital é moldado por disputas de poder e interesses específicos, desafiando a noção de uma democracia digital autêntica.

uma consulta em um mecanismo de pesquisa produzirá as informações mais relevantes e, portanto, úteis, quando, na verdade, ela é baseada em uma matriz de maneiras pelas quais as páginas são hiperlinkadas e indexadas na *web*.

As práticas algorítmicas do Google de enviesar informações em direção aos interesses do poder hegemônico, ao mesmo tempo em que apresentaram seus resultados como gerados a partir de fatores objetivos, resultaram em um fornecimento de informações que perpetua as caracterizações de mulheres e garotas por meio de *sites* misóginos e pornográficos. As próprias tecnologias de pesquisa e seu design não ditam ideologias raciais. Em vez disso, elas refletem e reinstanciam os valores sociais e culturais predominantes na sociedade.

O racismo algorítmico é uma prática que atualiza esse discurso sobre como as mulheres negras são subalternizadas. Noble (2013) identificou, em suas pesquisas de 2011, as formas hegemônicas nas quais identidades racializadas e de gênero eram retratadas e legitimadas por meio da pesquisa do Google, que, ao codificar a pornografia em primeiro plano como o tipo de informação mais importante ou significativa sobre mulheres negras, contribuiu para que esse discurso racializado e estereotipado se tornasse mais significativo, construindo a noção da mulher negra fetichizada.

Há, portanto, uma prática discursiva dessas categorias, “negra” e “mulheres/garotas”, moldada por relações de poder que objetivam normalizar e normatizar as formas de referenciar essas categorias. Assim, discursos hegemônicos sobre a mulher negra hipersexualizada, que já circulavam através de diferentes gêneros discursivos em meios não digitais, também se propagam on-line.

Ao comparar os resultados e os anúncios na busca por “garotas negras” com discursos mais amplos sobre mulheres e garotas negras, é possível observar a forma como a tecnologia do mecanismo de busca replica e instancia noções depreciativas. Esses discursos incluem discursos sobre mulheres negras como uma série de estereótipos históricos e modernos, como Jezebel, Sapphire e Mammy[26]. Durante a

[26] Segundo Collins (2019), as representações femininas negras como Jezebel, Sapphire e Mammy funcionam como símbolos de controle que reforçam os papéis sociais e as limitações impostas às mulheres negras, operando na cultura para reforçar a dominação racial, de gênero e de classe. Jezebel é o estereótipo da mulher negra como altamente sexualizada e promíscua, supostamente capaz de usar seu poder de sedução para enganar e manipular. A origem do termo vem de uma personagem bíblica que, na história, foi usada para justificar ataques sexuais de homens brancos às mulheres negras escravizadas. Na modernidade, a figura da Jezebel mantém essa conotação de sexualidade exagerada, alimentando uma visão da mulher negra como sexualmente perigosa e descontrolada.

escravidão, os estereótipos eram usados para justificar a vitimização sexual de mulheres negras por seus proprietários, dado que, segundo a lei, as mulheres negras eram propriedade e, portanto, não podiam ser consideradas vítimas de estupro.

Nesse sentido, a fabricação do estereótipo de Jezebel desempenhou um papel marcante ao retratar mulheres negras como sexualmente insaciáveis e gratuitas. De igual modo, os produtores de *sites*, pelas palavras ou pelos texto escolhidos, nas descrições de frases e anúncios, nos otimizadores de mecanismos de busca e nos conglomerados de mídia, reforçam um discurso colonial de subalternidade, pela exploração desses estereótipos que permanecem ativos nos processos algorítmicos atuais, sendo que os exemplos elencados ao longo deste estudo não deixam margem para uma interpretação que questione essa afirmativa.

Sapphire é a figura da mulher negra hostil, dominante, assertiva, agressiva e ameaçadora, que castra a masculinidade e atua de forma desafiante às normativas de feminilidade branca. Essa imagem associa a mulher negra a características de rudeza, fala altiva e comportamento provocador, que ameaça os padrões de passividade feminina considerados ideais na cultura branca hegemônica. Ela é vista como uma figura de desestabilização das normas sociais e de controle. Mammy, por sua vez, é o estereótipo da babá negra escravizada, obediente, abnegada e muitas vezes também como ama de leite dos bebês de uma família branca, retratada como uma figura maternal dedicada e dessexualizada, muitas vezes desconsiderada de seus desejos. É associada a uma imagem de respeito e cuidado, embora carregue uma carga de imposição de papéis de cuidado e submissão, além de ser uma figura que reforça a ideia de que as mulheres negras devem se dedicar ao trabalho doméstico de forma abnegada.

CONCLUSÕES

Esta pesquisa partiu da premissa de que os pressupostos levantados podem não ser comprovados de forma definitiva e que permanecem abertos a novas investigações, considerando a natureza dinâmica da análise do discurso, campo teórico que possibilita interpretações variadas, uma vez que os sujeitos que as produzem também possuem diferentes perspectivas. A abordagem focou, conforme estabelecido no objetivo geral da pesquisa, nos efeitos de sentido produzidos pelos algoritmos e nas formas como esses dispositivos atuam como mecanismos de biopoder, regulando e normatizando corpos racializados. Partiu-se do pressuposto de que a discriminação racial está embutida na programação de sistemas algorítmicos assentada nos mais diversos setores, podendo se tornar instrumento de controle, ampliar disparidades e exercer poder sobre populações estigmatizadas.

Compreender como as práticas discursivas sustentam o racismo algorítmico e se relacionam com desigualdades históricas e estruturais é abrir um campo de discussão, para explorar o papel que as tecnologias digitais têm no fortalecimento, na propagação e na perpetuação de estruturas de poder na sociedade. Como contribuição, esta pesquisa visou oferecer uma descrição atualizada sobre o racismo algorítmico, a partir de uma perspectiva foucaultiana do biopoder, um fenômeno emergente e complexo, que demanda uma reflexão ética e política sobre o papel dos algoritmos na vida contemporânea de cada sujeito.

Nesta pesquisa, não foram avaliados os registros de ocorrências de racismo algorítmico como produto da sociedade que o gerou de acordo com as relações de forças históricas, não localizadas no tempo e no espaço, mas dispersas, em que o poder é exercido. Mediante a discursividade presente na diversidade de ocorrências de racismo algorítmico elencadas ao longo desta pesquisa, o que se constitui um dos objetivos específicos da pesquisa – qual seja, mapear as representações sociais e os estereótipos sobre os grupos racializados –, tornou-se perceptível que há um racismo escamoteado, que ganha vida gradativa no meio social. O viés identificado nos discursos revela-se como um deslocamento histórico e social que guarda resquícios de uma memória cultural ativada por mecanismos discursivos que fazem o sujeito sujeitar-se a um padrão identitário, estereotipado nos diversos segmentos da sociedade.

Embora a IA e os algoritmos sejam ferramentas úteis para potencializar as capacidades humanas, eles apresentam o risco de reproduzir e agravar estigmas e preconceitos, tal como o racismo na sociedade brasileira. Assim, afirma-se neste trabalho não ser possível falar em neutralidade algorítmica, bem como a necessidade de se criar ferramentas jurídicas de enfrentamento, vigilância e punição ao racismo algorítmico. Mais do que denunciar a existência desse tipo de racismo, fato já fartamente documentado na literatura sobre o tema, esta pesquisa objetivou estabelecer um paralelo entre esse fenômeno e as dinâmicas racistas que permeiam as relações sociais ou, em outras palavras, compreender como as práticas discursivas que sustentam o racismo algorítmico se relacionam com desigualdades históricas e estruturais.

O percurso desta pesquisa levou à conclusão de o racismo algorítmico se constitui, não como um mero fenômeno oriundo do meio digital, mas sim como uma verdadeira formação discursiva, na esteira do é definido por Foucault (2019). Para se chegar a essa conclusão, há que se ressaltar a importância do método arqueológico proposto por esse pensador.

Nessa conjuntura, vale destacar que os algoritmos por si só não são bons nem ruins, não representam um mal a ser combatido, devido ao seu caráter racista, como um meio de dominação. O desafio da sociedade é modificar a forma como eles são construídos, para que seja apropriado e subvertido a favor do bem comum. Para isso, Benjamin aponta como caminho o combate à desigualdade codificada de modo mais eficaz e ativo, por meio de comunidades que ofereçam um conjunto de contradiscursos tecnológicos para combater o *New Jim Code* mediante a democratização dos dados (design e aplicação).

Nesse contexto, a apropriação tecnológica por grupos racializados a fim de combater o modelo discriminatório que reforça os impactos do racismo na sociedade é urgente e fundamental. Ao tornar visíveis as formas pelas quais os algoritmos participam na exclusão racial, pode-se começar a opor práticas as quais perenizam essas desigualdades. Percebe-se que a teoria foucaultiana, ao enfatizar a importância de desvelar os mecanismos de poder ocultos, oferece um guia para essa tarefa, incentivando a vigilância crítica e a resistência ativa. A detecção e a mitigação de vieses requerem não apenas avanços técnicos, mas também um entendimento profundo das implicações sociais e históricas. A revisão contínua de algoritmos e a inclusão de diversas perspectivas no processo de desenvolvimento são passos

essenciais para construir sistemas mais justos e equitativos, que estejam, obviamente, menos carregados de preconceitos.

Para que esse fim seja alcançado, a mobilização da sociedade é fundamental, à qual se soma o aparato jurídico cuja construção está em curso, mas que já possui um marco legal, com o arcabouço jurídico brasileiro de proteção de dados e da internet, notadamente a Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018) e o Marco Civil da Internet (Lei nº 12.965/2014), que deveriam constituir as bases fundamentais para proteção de usuários das ferramentas digitais, mas que são incapazes de enfrentar os desafios emergentes colocados pela inteligência artificial e mitigar eficazmente os vieses discriminatórios inerentes a esses sistemas automatizados.

A LGPD representou um avanço monumental ao espelhar o Regulamento Geral de Proteção de Dados (RGPD) europeu. Ela consagra princípios cruciais para o debate, como finalidade, adequação, transparência e não discriminação (Art. 6º). Dois artigos são especialmente relevantes: o Art. 6º, que veda o tratamento de dados para fins discriminatórios ilícitos ou abusivos; e o Art. 20, que garante ao titular o direito de solicitar a revisão de decisões automatizadas.

O Marco Civil da Internet, por sua vez, estabeleceu, em 2014, princípios basilares para o uso da internet no Brasil, como a neutralidade da rede, a liberdade de expressão e a proteção à privacidade. Seu Art. 7º, inciso X, assegura a “não suspensão da conexão à internet, salvo por débito diretamente decorrente de sua utilização”. No entanto, sua redação é anterior à massificação de sistemas de IA complexos, não abordando diretamente a *accountability* das plataformas quanto aos seus algoritmos.

A aplicação desses dispositivos ao racismo algorítmico, no entanto, é desafiadora. A “explicação” de uma decisão automatizada, prevista no Art. 20, muitas vezes, se reduz a uma descrição técnica incompreensível para o cidadão comum, não elucidando o viés racial incorporado à lógica do algoritmo (PASQUALE, 2015). A transparência, portanto, não se traduz automaticamente em inteligibilidade ou justiça. Além disso, a LGPD foca no tratamento de *dados pessoais*, mas o racismo algorítmico frequentemente emerge de inferências e correlações feitas a partir de grandes conjuntos de dados anonimizados, operando em um nível macro e sistêmico que pode escapar ao escopo individualista da lei (ZUBOFF, 2021).

O dispositivo legal, ao proibir a discriminação com base em “origem racial” (Art. 5º, XLII, CF/88 e Art. 6º, X, LGPD), enfrenta a dificuldade de provar a intenção discriminatória em um código de computador, que opera sob a aparente neutralidade da objetividade matemática. Fica evidente que a regulação atual, embora essencial, é reativa e insuficiente. A LGPD não é uma regulação específica para IA. Ela carece de mecanismos robustos de *auditoria algorítmica* obrigatória e independente que verifiquem a existência de vieses discriminatórios. O Marco Civil, por sua vez, não impõe obrigações de transparência algorítmica às plataformas.

A solução carece de um olhar intersetorial e propostas legislativas complementares. Projetos como o projeto de Lei 21/20, que estabelece um marco legal para a IA no Brasil, buscam preencher essa lacuna, prevendo princípios como a “supervisão humana” e a “não discriminação”. A doutrina aponta que é crucial incorporar a perspectiva dos Direitos Humanos no design tecnológico (*Privacy by Design* e *Fairness by Design*) e promover a diversidade nas equipes de desenvolvimento de IA para que vieses sociais sejam identificados e corrigidos (BENJAMIN, 2019).

Logo, a LGPD e o Marco Civil da Internet são avanços legais fundamentais que fornecem os alicerces para a defesa da privacidade e da autonomia do sujeito na era digital. No entanto, diante do desafio multifacetado do racismo algorítmico, mostram-se instrumentos limitados. Eles combatem a discriminação no nível do tratamento individual de dados, mas são menos eficazes contra a discriminação sistêmica e emergente embutida na lógica operacional de sistemas de IA.

Por isso, enfrentar de modo efetivo o racismo algorítmico exige ir além da proteção de dados. Impõe a criação de um arcabouço regulatório específico e proativo que obrigue à auditoria, à transparência significativa e à *accountability* dos algoritmos. Requer um compromisso ético com a equidade, que reconheça que a tecnologia não é um espaço neutro, mas um reflexo e um potencial amplificador das desigualdades e dos preconceitos existentes na sociedade. Assim sendo, a lei deve ser uma ferramenta para garantir que a inteligência artificial sirva à humanidade de forma justa e igualitária, e não para cristalizar hierarquias sociais sob um novo disfarce tecnológico.

REFERÊNCIAS

ALAGIĆ, A. *et al.* Application of artificial intelligence in the analysis of the facial skin health condition. *IFAC-PapersOnLine*, 2022. v.55, n.4, p.31-37. Disponível em: <https://doi.org/10.1016/j.ifacol.2022.06.005>.

ALLPORT, G. W. *The nature of prejudice*. 3.ed. Wokingham: Addison-Wesley, 1954.

AMODIO, D. M. The neuroscience of prejudice and stereotyping. *Nature Reviews Neuroscience*, 2014.

ARROYO, C. L. Constitución, Derechos Fundamentales, Inteligencia Artificial Y Algoritmos. *Revista do Direito*, 2022, n.66, p.139-158.

ÁVILA-TOMÁS, J. F.; MAYER-PUJADAS, M. A.; QUESADA-VARELA, V. J. Artificial intelligence and its applications in medicine II: Current importance and practical applications. *Atencion Primaria*, 2020, v.53, n.1, p.81-88.

BAROCAS, Solon; SELBST, Andrew D. *Algorithmic Bias Detectable and Mitigable*: Best Practices and Policies to Reduce Consumer Harms. 2023. In: https://www.studocu.com/row/document/university-of-bahrain/artificial-intelligence/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms-brookings/91627592. Acessado em 23/08/2024.

BARONAS, Roberto Leiser. *Encontros e desencontros epistemológicos entre Foucault e Pêcheux*: Formação Discursiva. 2011. Trabalho apresentado no V Seminário de Estudos em Análise do Discurso, Porto Alegre, pp.20-23 set. 2011.

BENJAMIN, Ruha. *Race After Technology*: Abolitionist Tools for the New Jim Code. Oxford: Oxford University Press, 2019.

BENJAMIN, Ruha. *Race after technology*: abolitionist tools for the New Jim Code. New York: Polity, 2019.

BENJAMIN, Ruha. Engaging Science. *Technology, and Society 2*, 2016, p.145-156.

BENJAMIN, Ruha. Retomando nosso fôlego: Estudos de Ciência e Tecnologia, Teoria Racial Crítica e a imaginação carcerária. In: SILVA, Tarcízio (Org.). *Comunidades, algoritmos e ativismos digitais*: Olhares afrodiaspóricos. Trad. Vinícius Silva, Tarcízio Silva. São Paulo: LiteraRUA, 2021. pp.13-26.

BENTO, Maria Aparecida Silva. Branqueamento e branquitude no Brasil. CARONE Iray; BENTO, Maria Aparecida Silva (Orgs.). *Psicologia social do racismo*: estudos sobre branquitude e branqueamento no Brasil. 4.ed. Petrópolis, RJ: Vozes, 2009. pp.25-57.

BEZERRA, Arthur Coelho *et al*. Pele negra, algoritmos brancos: informação e racismo nas redes sociotécnicas. *Liinc em Revista*, v.18, n.2, p. e6043-e6043, 2022.

BHBOSALE, S.; PUJARI, V.; MULTANI, Z. Advantages and disadvantages of artificial intelligence. *Aayushi International Interdisciplinary Research Journal*, 2020, v.77, n.October, p.227–230. Disponível em: <https://towardsdatascience.com/advantages-and-disadvantages-of-artificial-intelligence-182a5ef6588c>.

BONILLA-SILVA, E. *Racismo sem Raças*: Negação e Invisibilidade no Brasil. Rio de Janeiro: Vozes. 2003.

BOONIPAT, T. *et al*. Using artificial intelligence to analyze emotion and facial action units following facial rejuvenation surgery. *Journal of Plastic, Reconstructive and Aesthetic Surgery*, 2022, n.XXV, p.8-10.

BROWNE, S. *Dark Matters*: On the Surveillance of Blackness. Durham: Duke University Press, 2015.

BUOLAMWINI, Joy Adowaa. Gender Shades: Intersectional Phenotypic and Demographic Evaluation of Face Datasets and Gender Classifiers. *B.S. in Computer Science, Georgia Institute of Technology*, 2017. Massachusetts Institute of Technology.

BUOLAMWINI, Joy & GEBRU, Timnit. Gender shades: Intersectional accuracy disparities in commercial gender classifcation. In Conference on fairness, accountability and transparency. *Proceedings of Machine Learning Research*, v.81, pp.1-15, 2018. In: http://proceedings.mlr. press/v81/buolamwini18a/buolamwini18a.pdf. Acessado em 20 jun.

BURAWOY, Michael. *A outra face da diferença*: uma antropologia da desigualdade. Trad. Sérgio Tadeu de Camargo. São Paulo: Editora Unesp, 2008.

CARNEIRO, Aparecida Sueli. *A construção do outro como não-ser como fundamento do ser*. 2005. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2005.

CARPENTER, K. A.; HUANG, X. Machine learning-based virtual screening and its applications to Alzheimer's drug discovery: a review. *Current pharmaceutical design*, 2018, v.24, n.28, pp.3347-3358.

CARRERA, Fernanda. Algoritmização de estereótipos raciais em bancos de imagens: a persistência dos padrões coloniais Jezebel, Mammy e Sapphire para mulheres negras. *Palabra Clave*, v.24, n.3, 2021.

COIMBRA, Jéssica Pérola Melo *et al*. Interseções Entre Racismo Algorítmico, Reconhecimento Facial e Segurança Pública no Brasil. *Revista Jurídica do Cesupa*, v.4, n.2, p.136-160, 2023.

COLLINS, Patricia Hill. *Pensamento feminista negro*: conhecimento, consciência e a política do empoderamento. São Paulo: Boitempo, 2019.

CORMEN, Thomas H. *et al*. *Algoritmos*: Teoria e Prática. 3.ed. Rio de Janeiro: Elsevier, 2012.

CORREIA, I.; BRITO, R.; VALA, J. & PÉREZ, J. *Normes Antiracistes et Persistance du Racisme Flagrant*: analyse comparative des attitudes face aux Tziganes et face aux noirs au Portugal. Manuscrito não-publicado. Centro de Investigação e Intervenção Social/ISCTE, 2001.

COZMAN, Fabio Gagliardi; KAUFMAN, Dora. Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação. *Revista USP*, n.135, p.195-210, 2022.

CRAWFORD, K. *Atlas of AI*: Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven: Yale University Press, 2021.

DA ROCHA, Cláudio Jannotti; PORTO, Lorena Vasconcelos; ABAURRE, Helena Emerick. *Discriminação algorítmica no trabalho digital*. *Revista de Direitos Humanos e Desenvolvimento Social*, v.1, pp.1-21, 2020.

DA SILVA, Mozart Linhares; ARAÚJO, Willian Fernandes. Biopolítica, racismo estrutural-algorítmico e subjetividade. *Educação Unisinos*, v.24, n.1, pp.1-20, 2020a.

DA SILVA, Tarcízio. 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.

DELGADO, R. & STEFANCIC, J. *Critical Race Theory*: An Introduction. New York: NYU Press, 2017.

DOMINGOS, Pedro. *O Algoritmo Mestre*. Como a busca pelo Algoritmo de Machine Learning definitivo recriará nosso mundo. São Paulo: Novatec, 2017.

DOVIDIO, J. F. On the nature of contemporary prejudice: the third wave. *Journal of Social Issues*, 57, pp.829-849, 2001.

DRUMOND, Erick Soares. *O livro didático e os privilégios da branquitude na formação de professores de Língua Inglesa*. Dissertação (Mestrado em Letras) - Instituto de Ciências Humanas e Sociais. Ouro Preto: Universidade Federal de Ouro Preto, Mariana, 2021.

DUCKITT, J. H. *The social psychology of prejudice*. Westport: Praeger Publishers/Greenwood Publishing Group, 1992.

ELIAS, P. S. Algoritmos, Inteligência Artificial e o Direito. *ConJur*, 2017. Disponível em: <https://www.conjur.com.br/dl/algoritmos-inteligencia-artificial.pdf>. Acesso em: 3 out. 2022.

EUBANKS, Virginia. *Automatizando a desigualdade*: como as ferramentas de alta tecnologia perfilam, policiam e punem os pobres. Trad. Sérgio Tadeu de Camargo. São Paulo: Editora Unesp, 2018.

FANON, Frantz. *Pele negra, máscaras brancas*. Salvador: EDUFBA, 2008.

FANON, F. Racismo e cultura. *Revista Convergência Crítica*, n.13, pp.78-90, 2018. Disponível em: https://periodicos.uff.br/convergenciacritica/article/view/38512. Acesso em: 05 jan. 2024.

FERNANDES, F. *A integração do negro na sociedade de classes (o legado da "raça branca")*. 5.ed. São Paulo: Globo, 2008.

FERNANDES, F. *O negro no mundo dos brancos*. 2.ed. São Paulo: Cortez, 2007.

FERNANDES, Sérgio Rocha. *Simulacros de diversidade racial*: permanência da ideologia do branqueamento no Brasil. Dissertação (Mestrado em Comunicação e Semiótica) - Programa de Estudos Pós-Graduados em Comunicação e Semiótica da Pontifícia Universidade Católica de São Paulo, São Paulo, 2023.

FOUCAULT, Michel. *O Nascimento da Clínica*. Rio de Janeiro: Forense Universitária, 1977.

FOUCAULT, Michel. *História da Loucura na Idade Clássica*. Trad. José Teixeira Coelho Neto. São Paulo: Perspectiva, 2017.

FOUCAULT, Michel. *História da sexualidade I*. Trad. Maria Thereza da Costa Albuquerque e J. A. Guilhon Albuquerque. Rio de Janeiro: Edições Graal, 1998.

FOUCAULT, Michel. *As palavras e as coisas*: uma arqueologia das ciências humanas. Trad. Salma Tannus Muchail. 8.ed. São Paulo: Martins Fontes, 2000.

FOUCAULT, Michel. *A Verdade e as Formas Jurídicas*. Trad. Alessandro A. de Assis. Rio de Janeiro: NAU, 2002.

FOUCAULT, Michel. *Segurança, Território, População*: curso dado no Collège de France (1977-1978). São Paulo: Martins Fontes, 2008.

FOUCAULT, Michel. *Estratégia, Poder-Saber*: Ditos e escritos. 2.ed. Vol. IV. Trad. Vera Lucia Avellar Ribeiro. Rio de Janeiro: Forense Universitária, 2010.

FOUCAULT, Michel. *A ordem do discurso*. Trad. Laura Fraga Sampaio. 23.ed. São Paulo: Edições Loyola, 2013.

FOUCAULT, Michel. *A Arqueologia do Saber*. 8.ed. Rio de Janeiro: Forense Universitária, 1969/2019.

FOUCAULT, Michel. *Vigiar e Punir*: Nascimento da Prisão. 42.ed. Petrópolis: Vozes, 2020.

FOUCAULT, Michel. *Microfísica do Poder*. Org. e trad. Roberto Machado. 11.ed. São Paulo: Paz e Terra, 2021.

FOUCAULT, Michel. *Em defesa da sociedade*. Trad. Maria Ermantina Galvão. São Paulo: Martins Fontes, 2021b.

GAERTNER, S. L. & Dovidio, J. F. The aversive form of racism. In: DOVIDIO, J. F. & GAERTNER, S. L. (Orgs.). *Prejudice, discrimination, and racism*. Orlando, Florida: Academic, 1986. pp.61-89.

GONZALEZ, L.; HASENBALG, C. *Lugar de negro*. Rio de Janeiro: Marco Zero, 1982.

GROSFOGUEL, Ramón. *Dilemas dos estudos étnicos norte-americanos*: multiculturalismo identitário, colonização disciplinar e epistemologias descoloniais. *Tábula Rasa*, Bogotá, Colômbia, p.17-48, n.4, enero-junio 2006.

HAN, Byung-Chul. *Infocracia*: digitalização e a crise da democracia. Trad. Gabriel S. Philipson. Petrópolis: Vozes, 2022.

HAN, Byung-Chul. *Sociedade do cansaço*. Trad. Enio Paulo Giachini. Petrópolis: Vozes, 2025.

HAYKIN, S. *Neural Networks and Learning Machines*. New Jersey: Prentice Hall, 2008.

HELMOND, ANNE. A plataformização da web. In: OMENA, J. J. *Métodos Digitais:* Teoria-Prática-Crítica. Lisboa: ICNOVA, 2019. pp.49-72.

HIMANSHU; KHANNA, R.; KUMAR, A. Artificial intelligence applications for target node positions in wireless sensor networks using single mobile anchor node. *Computers and Industrial Engineering*, 2022. v.167, n. February.

HOFBAUER, Andreas. O conceito de “Raça” e o ideário do branqueamento no século XIX – Bases ideológicas do racismo brasileiro. In: *Teoria e Pesquisa 42 e 43*. São Paulo: UFSCAR, 2003.

H, R. Artificial Intelligence based facial recognition for Mood Charting among men on life style modification and it’s correlation with cortisol. *Asian Journal of Psychiatry*, 2019 v.43, n.May, p.101-104. Disponível em: <https://doi.org/10.1016/j.ajp.2019.05.017>.

IBGE. Diretoria de Pesquisas, Coordenação de População e Indicadores Sociais, Pesquisa Nacional por Amostra de Domicílios Contínua 2018.

LEE, Nicol Turner; RESNICK; Paul & BARTON, Genie. *Detecção e mitigação de viés algorítmico*: melhores práticas e políticas para reduzir danos ao consumidor. https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/ 2019.

LIMA, M. E. O. & Vala, J. Individualismo meritocrático, diferenciação cultural e racismo. *Análise Social*, 37, pp.181-207, 2002.

MASTRODICASA, D. *et al*. Artificial Intelligence Applications in Aortic Dissection Imaging. *Seminars in Roentgenology*, 2022.

MAYER-SCHÖNBERGER, V. & CUKIER, K. Big data: A revolution that will transform how we live, work, and think. *Houghton Mifflin Harcourt*, 2002.

MBEMBE, Achille. *Crítica da Razão Negra*. Trad. Sebastião Nascimento. São Paulo: n-1 Edições, 2018.

MBEMBE, Achille. *Necropolítica*: biopoder, soberania, estado de exceção, política da morte. Trad. Renata Santini. São Paulo: n-1 Edições, 2018b.

MIRA, J. M. Symbols versus connections: 50 years of artificial intelligence. *Neurocomputing*, 2008. v.71, n.4-6, p.671-680.

MOACIR, R. F. de M. & PONTI, M. A. *Machine learning A Practical Approach on the Statistical Learning Theory*. [S.l.]: [s.n.], 2017. v.45.

MUNANGA, Kabengele. *Rediscutindo a mestiçagem no Brasil*: identidade nacional versus identidade negra. 3.ed. Belo Horizonte: Autêntica, 1999.

NAKAMURA, L. & CHOW-WHITE, P. Introduction - Race and Digital Technology: Code, the Color Line and Information Society. In: NAKAMURA, L. CHOW-WHITE, P. (Orgs.). *Race after the internet*. New York: Routledge, 2012, pp.1-18.

NOBLE, Safiya Umoja. Google Search: Hyper-visibility as a Means of Rendering Black Women and Girls Invisible. *Invisible Culture Journal*, n.19, [oct. 2013]. Disponível em:<https://www.invisibleculturejournal.com/pub/google-search-hypervisibility/release/1?readingCollection=cf7d0642>. Acesso em: [06/12/2024].

NOBLE, Safiya Umoja. *Algoritmos da opressão*: como o Google fomenta e lucra com o racismo. Santo André, SP: Rua do Sabão, 2021.

NOBRE, M. *Inteligência Artificial e Democracia*: Desafios para a Regulação. São Paulo: Companhia das Letras, 2021.

NORIEGA, M. The application of artificial intelligence in police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, cooperation, and false confessions. *Futures*, 2020. v.117, p.102510. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0016328719303726>.

OLIVEIRA, Rafael Paraguassu de. Score de Crédito e o Algoritmo (inteligência artificial) utilizado pelas Instituições Financeiras: impacto, responsabilidades e discriminações. *Revista Cognitio Juris*, ano XIII, n.48, julho de 2023. Disponível em: https://cognitiojuris.com.br/score-de-credito-e-o-algoritmo-inteligencia-artificial-utilizado-pelas-instituicoes-financeiras-impacto-responsabilidades-e-discriminacoes/#_ftn1.

O'NEIL, Cathy. *Algoritmos de destruição em massa*: como o Big Data aumenta a desigualdade e ameaça à democracia. Trad. Rafael Abraham. Santo André, SP: Rua do Sabão, 2020.

ORLANDI, Eni P. *Análise de Discurso*: Princípios e Procedimentos. Campinas: Pontes Editores, 2009.

PASQUALE, F. *The Black Box Society*: The Secret Algorithms That Control Money and Information. Cambridge: Harvard University Press, 2015.

PÊCHEUX, Michel. Análise automática do discurso (AAD-69). In: GADET, F. & HAK, T. (Orgs.). *Por uma Análise Automática do Discurso*. Campinas: Unicamp, 2010. pp.59-158.

PÊCHEUX, Michel. *Semântica e Discurso*: uma crítica à afirmação do óbvio. São Paulo: UNICAMP, 2014.

PEIXOTO, F. H. Direito E Inteligência Artificial Na (Não) Redução De Desigualdades Globais: Decisões Automatizadas Na Imigração E Sistemas De Refugiados. *Revista Direitos Culturais*, 2020, v.15, n.37, p.305-320.

PETTIGREW, T. F. Personality and social-cultural factors in intergroup attitudes: a cross-national comparison. *Journal of Conflict Resolution*, 2, pp.29-42, 1958.

PETTIGREW, T. F. Regional differences in antinegro prejudice. In: ARONSON, E. & PRATKANIS, A. R. (Orgs.). *Social Psychology III*. Brookfield: Edward Elgar, 1993. pp.359-367.

PIOVEZANI, Carlos. *Análise do Discurso*: Sujeito, Sentido e Ideologia. São Paulo: Contexto, 2015.

RAMOS, A. G. *A Discriminação no Brasil*. Rio de Janeiro: Zahar Editores, 1978.

RAMOS, A. G. *Introdução crítica à Sociologia brasileira*. Rio de Janeiro: Editora da UFRJ, 1995.

REIS, D. S. Saberes encruzilhados: (de) colonialidade, racismo epistêmico e ensino de filosofia. *Educar em Revista*, Curitiba, v.36, e75102, 2020.

REZENDE, Milka de Oliveira. Racismo no Brasil (2025). In: https://mundoeducacao.uol.com.br/sociologia/racismo-no-brasil.htm. Acessado em: 12 jan. 2025.

SALDANHA, Felipe Gustavo Guimarães. *Jornalismo escolar educomunicativo*: uso de linguagens e procedimentos midiáticos pelo Programa Imprensa Jovem da rede municipal de educação de São Paulo. Tese (Doutorado). São Paulo: Universidade de São Paulo, 2023.

SANTOS, Boaventura de Sousa. *A Epistemologia do Sul*. São Paulo: Cortez, 2010.

SARLET, Ingo Wolfgang. *Inteligência Artificial, Proteção de Dados Pessoais e Responsabilidade na Era Digital*. São Paulo: Saraiva Educação SA, 2022. Série Direito, Tecnologia, Inovação e Proteção de Dados num Mundo em Transformação.

SHAPIRO, S. C. *Encyclopedia of artificial intelligence second edition*. New Jersey: A Wiley Interscience Publication, 1992. p.1-9.

SILVA, Lucas Vinicius Nunes & GONÇALVES, Cristhóvão Fonseca. *Perspectivas sobre racismo estrutural e tráfico de drogas*: a política de filtragem racial na atuação policial do estado de Pernambuco. Ano 8, n.6, 2022.

SILVA, R. L. Da; SILVA, F. dos S. R. da. Reconhecimento facial e segurança pública: os perigos do uso da tecnologia no sistema penal seletivo brasileiro. In: Congresso Internacional de Direito e Contemporaneidade, Santa Maria, RS, Brasil. 2019.

SILVA, T. Colonialidade difusa no aprendizado de máquina: camadas de opacidade algorítmica na internet. In: SILVEIRA, S. A. da; SOUZA, J. & CASSINO, J. F. (Orgs.). *Colonialismo de Dados*. São Paulo: Autonomia Literária, 2021. pp.87-108.

SILVA, T. *Racismo algorítmico*: inteligência artificial e discriminação nas redes digitais. Edições Sesc SP, 2022a. E-Pub, II (Democracia Digital)

SILVA, T. Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código. In: SILVA, Tarcízio (Org.). *Comunidades, Algoritmos e Ativismo Digitais*: olhares afrodiaspóricos. São Paulo: LiteraRUA, 2020.

SILVA, T. *Teoria racial crítica e comunicação digital*: conexões contra a dupla opacidade. Artigo apresentado no 42º Congresso Brasileiro de Ciências da Comunicação. Belém, PA, Brasil, 2019a. https://www.bit.ly/3ju6KUb

SILVA, T. Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código. *VI Simpósio Internacional LAVITS – Assimetrias e (In)Visibilidades*: vigilância, Gênero e Raça, 2019b. pp.1-17.

SILVA, T. Mapeamento de Danos e Discriminação Algorítmica. *Desvelar*, 2023. Disponível em: https://desvelar.org/casos-de-discriminacao-algoritmica/. Acesso em: 30/10/2024.

SILVEIRA, S. A.; MOURA, L. do V.; ALMEIDA, L. T. G. de. A reprogramação da sociedade nos discursos sobre algoritmos. *Surveillance in Latin America*, 2019. p.1-22. Disponível em: <https://lavits.org/eventos/simposio-lavits-2019/?lang=en>.

SOUSA, Vinicius Dino de. *O problema do algorithmic bias (viés algorítmico) no auxílio aos juízes de Direito pela inteligência artificial*: Uma investigação sobre a imparcialidade e injustiça da inteligência artificial. 2020, https://www.jusbrasil.com.br/artigos/o-problema-do-algorithmic-bias-vies-algoritmico-no-auxilio-aos-juizes-de-direito-pela-inteligencia-artificial/825348884. Acessado em 15/08/2024.

SRNICEK, Nick. *Platform capitalism*. Cambridge: Polity Press, 2017.

SWEENEY, Latanya. *Discrimination in Online Ad Delivery* (28 de janeiro de 2013). Disponível em SSRN: https://ssrn.com/abstract=2208240 ou http://dx.doi.org/10.2139/ssrn.2208240.

TIM WU. *Network Neutrality, Broadband Discrimination*. 2 J. on Telecomm. & High Tech. L. 141 (2003).

TRIVEDI, R. & KHADEM, S. Implementation of artificial intelligence techniques in microgrid control environment: Current progress and future scopes. *Energy and AI*, 2022, v.8, n. March, p.1-19.

TSIKTSIRIS, D. *et al*. A Novel Image and Audio-based Artificial Intelligence Service for Security Applications in Autonomous Vehicles. *Transportation Research Procedia*, 2022. v.62, n. Ewgt 2021, pp.294-301.

TUFEKCI, Z. Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency. *Colorado Technology Law Journal, 13,* pp.203-218, 2015.

VAN DIJK, J. *The Digital Divide*: The Internet and Social Inequality in International Perspective. SAGE Publications, 2020.

VILELA, Eugênia. Do Biopoder: Ensaio sobre a relação entre a política e a vida no pensamento de Michel Foucault. *Revista de Estudos Universitários – REU*, v.37, n.2, 2011.

WHITTAKER, M.; CRAWFORD, K. *Discriminating Systems*: Gender, Race, and Power in AI. [s.l.]: AI Now Institute, 2019. Disponível em: https://ainowinstitute.org/. Acesso em: 21 jan. 2024.

ZHOU, J. *et al*. Application of artificial intelligence in the diagnosis and prognostic prediction of ovarian cancer. *Computers in Biology and Medicine*, 2022, v.146, n.February, p.105608. Disponível em: <https://doi.org/10.1016/j.compbiomed.2022.105608>.

ZHU, L. *et al*. Can artificial intelligence enable the government to respond more effectively to major public health emergencies? Taking the prevention and control of Covid-19 in China as an example. *Socio-Economic Planning Sciences*, 2022, v.80, n. December 2020, pp.1-9.

ZUBOFF, S. *A era do capitalismo de vigilância*. Rio de Janeiro: Intrínseca, 2021.

ANEXOS

Anexo I – Ocorrência 1

Online

TECNOLOGIA

Racismo algorítmico: a exclusão da população negra

Inteligência artificial e algoritmos estão surgindo e se consolidando em pleno vapor, reproduzindo o racismo estrutural presente na sociedade

Liliane Rocha (21 de novembro de 2023)

Há poucas semanas, repercutiu nas redes sociais uma *trend* da qual muitos influenciadores, formadores de opinião, empresários, entre outros, participaram. A *trend* da Disney Pixar, em que, com o apoio da inteligência artificial, era possível criar imagens semelhantes a nós, remetendo aos personagens dos desenhos animados da Pixar. Pois bem, na primeira tentativa para criar a minha "personagem Pixar", o app em questão pediu uma foto, e minha equipe de comunicação fez o upload. E, surpresa, não fosse o fato de que a imagem criada era de uma pessoa branca, o desenho veio igualzinho a mim.

Na segunda tentativa, a equipe mais uma vez entrou na plataforma indicada pelos sites de tecnologia, escreveu todos os parâmetros, e *voilà*, como diriam os franceses, novamente uma imagem de pessoa branca. Para que o resultado fosse a imagem de uma pessoa negra, era preciso especificar: executiva negra. Ao colocar somente a palavra executiva, a imagem gerada era compulsoriamente de uma pessoa branca. Não à toa, por curiosidade, fui olhar a página de algumas personalidades negras nas redes sociais e nos desenhos criados para essa *trend*: não de todos, mas de muitos, a pele do personagem também estava mais clara.

A tentativa bem-sucedida de Liliane Rocha, só depois de escrever "executiva negra"
(Reprodução/Instagram/@lilianerochaoficial)

Aliás, não sei se você já prestou atenção, mas no Instagram, quando vamos fazer um stories, e tentamos inserir um ícone, se escrevemos a palavra anjo, virá sempre a imagem de um anjo branco.

Atleta, virão sempre atletas brancos. E assim por diante. Para vir uma imagem de anjo negro, atleta negro, ou qualquer outra iconização negra, eu tenho que especificar a palavra negro ao final. Como se a branquitude fosse norma e padrão, e as outras etnias necessitassem ou demandassem ser detalhadas como uma exceção.

O mesmo acontece sempre que entramos em aplicativos de tratamentos de fotos. Os traços afrodescendentes são suavizados, nariz e boca levemente afinados, tom da pele sutilmente clareado, contornos levemente retirados. Porém, de sutileza em sutileza se delineia o racismo. Exatamente como o app que, para teoricamente nos embelezar, subtrai os traços étnicos raciais negros.

Enfim, o assunto não é novo, muito menos desconhecido da grande maioria. Aliás, talvez esse seja o problema, tendo em vista do tanto que o assunto está velho e batido e, também o quanto é relevante dado o advento dos apps e plataformas. Nem sabemos ao certo como essas novas tecnologias vão impactar as nossas vidas. Por isso, os algoritmos serem racistas é grave. Aliás, gravíssimo.

Podemos dizer que a IA e os algoritmos estão surgindo e se consolidando em pleno vapor, reproduzindo o racismo estrutural presente na sociedade, nas empresas e no Vale do Silício.

Há, ainda, mais uma face dessa problemática. Neste mês, no qual prioritariamente falamos sobre Consciência Negra, tentamos impulsionar em nosso perfil do Instagram duas artes e textos, com os seguintes conteúdos: “Se somos um país miscigenado, então como opera o racismo?”, e “Seríamos nós todas iguais? Lutamos todas pelos mesmos ideais?”. Ambos muito relevantes e com análise de legislação, pesquisas e afins. Sem nenhuma menção político-partidária, ou ofensa de qualquer tipo, mesmo assim os conteúdos foram reprovados pela plataforma, sem motivo ou justificativa. O retorno que obtivemos foi de que não estavam dentro da política. Simples assim!

Não à toa, influenciadores negros têm recorrentemente chamado a atenção para o fato de terem muito menos repercussão em suas postagens e nas entregas do seu conteúdo quando comparados com os influenciadores brancos. Mesmo se a foto, contexto e anúncio for extremamente semelhante. Segundo o site Negrê, “a digital influencer e youtuber Sá Ollebar, criadora do projeto digital Preta Pariu, iniciou um experimento na plataforma Instagram (…) Após perceber a crescente queda nos índices de alcance digital, a paulista publicou fotografias de modelos caucasianas (brancas) em seu perfil e analisou as métricas de engajamento. Surpreendentemente, a ferramenta de estatísticas aferiu um aumento de 6000% em seu alcance”.

Parafraseando e completando a pergunta que ela fez, o Instagram, e as demais redes sociais, algoritmos e as IAs, só nos representam, entregam nosso conteúdo e nos contemplam se formos brancos?

No médio e longo prazo, qual será o impacto desse racismo algorítmico que estamos deixando passar em vão na sociedade? Precisamos, urgentemente, de mais programadores negros, negras, LGBTQPIAN+, com deficiência, mulheres, periféricos e todas as ramificações da diversidade também trabalhando na concepção e programação dessas ferramentas.

Como tenho dito, a realidade só muda quando estamos construindo, produzindo e pensando juntos! E mais, também nos cargos de tomada de decisão. Seguindo essa lógica, já faço este texto antevendo que será bloqueado pelos algoritmos e lido por poucos. Ainda assim, insisto, pois o assunto

é necessário e urgente. Conto com você, para juntos furarmos essa bolha. Cobre as empresas de tecnologia, contrate pessoas negras, compartilhe este artigo. Quem sabe assim, podemos fazer com que a era da Inteligência Artificial seja a primeira grande revolução da história da humanidade realmente inclusiva.

Liliane Rocha é mestre em Políticas Públicas, CEO e Fundadora da Gestão Kairós, consultoria de diversidade e sustentabilidade.

Anexo II – Ocorrência 2

EL PAÍS

Tecnologia

Aplicativo FaceApp 'branqueia' os usuários para torná-los "mais sexy"

Surgem críticas porque o filtro para parecer mais 'sensual' clareia o tom da pele e elimina os óculos

El País

Madri - 25 ABR 2017 - 18:13 BRT

O efeito do filtro do FaceApp para se ver mais 'sexy'. *Twitter*.

O aplicativo FaceApp, que permite ao usuário modificar fotos do rosto para ficar sorridente ou com um aspecto envelhecido, foi acusado de racismo porque seu filtro para parecer mais *sexy* branqueia a pele. Os criadores do *app* para celulares enviaram um pedido de desculpas pelo que consideraram "um efeito secundário infeliz" da tecnologia utilizada nos filtros.

Vários usuários negros postaram nas redes sociais imagens que mostram como esse filtro branqueou a pele deles. "Então, baixei esse *app* e decidi utilizar o filtro sexy sem saber que me tornaria branco. É 2017, por favor, pessoal!", escreveu Shahquelle L., um rapaz de 21 anos. "#faceapp não só é ruim, é também racista. O filtro sexy = branquear minha pele e fazer de meu nariz sua opinião do que é europeu. Não, obrigado, #desinstalado", se queixou Terrance AB Johnson, de Seattle, EUA.

O aplicativo, lançado em janeiro, oferece várias possibilidades para modificar o rosto. Uma opção é colocar um sorriso em quem está sério, outra permite saber como um homem seria se fosse mulher, ou uma mulher se fosse homem. Também há filtros que rejuvenescem ou envelhecem quem aparece na tela.

Uma usuária canadense que testou vários desses filtros constatou que a opção para parecer mais sexy lhe tirava os óculos e também tornava mais pálida a sua pele, que já é branca. “Simplesmente esquisito”, avaliou em sua conta do Twitter, @Caitofthenorth.

A rede britânica BBC contatou o executivo-chefe e fundador da empresa que lançou o aplicativo, Yaroslav Goncharov, que pediu desculpas. “Estamos profundamente arrependidos por esta questão inquestionavelmente grave”, escreveu em um comunicado. O desenvolvedor afirmou se tratar de “um efeito secundário infeliz da rede neural”, o tipo de inteligência artificial que o aplicativo utiliza para remodelar os rostos fundindo diferentes características faciais.

O FaceApp decidiu renomear o efeito, que passou de *hot (sinônimo de sexy)* para *spark* (faísca), mas continua aplicando o branqueamento da pele.

O aplicativo, lançado em janeiro, oferece várias possibilidades para modificar o rosto. Uma opção é colocar um sorriso em quem está sério, outra permite saber como um homem seria se fosse mulher, ou uma mulher se fosse homem. Também há filtros que rejuvenescem ou envelhecem quem aparece na tela.

Uma usuária canadense que testou vários desses filtros constatou que a opção para parecer mais sexy lhe tirava os óculos e também tornava mais pálida a sua pele, que já é branca. “Simplesmente esquisito”, avaliou em sua conta do Twitter, @Caitofthenorth.

A rede britânica BBC contatou o executivo-chefe e fundador da empresa que lançou o aplicativo, Yaroslav Goncharov, que pediu desculpas. “Estamos profundamente arrependidos por esta questão inquestionavelmente grave”, escreveu em um comunicado. O desenvolvedor afirmou se tratar

de “um efeito secundário infeliz da rede neural”, o tipo de inteligência artificial que o aplicativo utiliza para remodelar os rostos fundindo diferentes características faciais.

O FaceApp decidiu renomear o efeito, que passou de *hot (sinônimo de sexy)* para *spark* (faísca), mas continua aplicando o branqueamento da pele.

Anexo III – Ocorrência 3

PODER 360

Prefeito de SP vende reconhecimento facial como solução mágica

Sem discussão, cidade vai instalar 20.000 câmeras de segurança que acumulam problemas de falso reconhecimento e racismo, escreve Mario Cesar Carvalho.

Instalação de câmeras na cidade é parte do programa Smart Sampa

Mario Cesar Carvalho 9.ago.2023 (quarta-feira) - 6h00

Políticos adoram vender soluções mágicas para problemas complexos. Principalmente se a mágica vier embalada com a aura digital, de novíssima tecnologia. Eles sabem que o problema não será resolvido, mas o que importa é a imagem de que algo muito moderno está sendo feito.

O prefeito de São Paulo, Ricardo Nunes (MDB), elegeu o reconhecimento facial para funcionar como a poção que vai solucionar uma questão que nem as democracias mais ricas do mundo conseguiram resolver – o tráfico de drogas, sobretudo daquelas que produzem imagens degradantes nas ruas, como o crack no Brasil, os opioides nos Estados Unidos e a heroína na Europa.

Nunes assinou na última 2ª feira (7.ago.2023) um contrato que prevê a instalação de 20.000 câmeras até o próximo ano, quando ele deve disputar a reeleição. O programa, chamado de Smart Sampa, custará R$ 9,2 milhões ao mês para a cidade.

As primeiras 200 câmeras serão instaladas na região conhecida como Cracolândia com o objetivo de combater o tráfico.

Reconhecimento facial é um campo minado. As câmeras e sistemas de inteligência artificial que executam essa tarefa estão sendo banidas de uma série de cidades nos Estados Unidos. A próxima da lista deve ser Boston, centro de alta tecnologia na costa leste e sede de algumas das mais importantes universidades do mundo, como Harvard e MIT (Massachusetts Institute of Technology).

Nova York e São Francisco, o epicentro do Vale do Silício, também proíbem o uso de reconhecimento facial para fins criminais.

A União Europeia aprovou em maio a 1ª versão de um projeto de lei que prevê o banimento do reconhecimento facial em espaços públicos. Houve uma grita enorme: muitos países querem que haja ao menos duas exceções, em casos de terrorismo ou de ameaça à segurança nacional.

Curiosamente, o prefeito de São Paulo se alinha com uma ditadura ao adotar o reconhecimento facial: é na China que essa tecnologia é mais usada.

Os argumentos usados para o banimento são diferentes nos Estados Unidos e na União Europeia. Enquanto cidades e Estados norte-americanos frisam a quantidade brutal de erros que esses sistemas incorrem, os parlamentares do bloco europeu argumentam que a tecnologia viola um bem com o qual não se pode negociar: a privacidade.

Os 2 argumentos foram usados contra o projeto do prefeito de São Paulo, mas o projeto será implantado praticamente sem discussão: houve um único dia de audiência pública, no qual os questionamentos ao projeto foram refutados com a habitual má vontade do poder público quando se questiona por que priorizam um projeto sobre o qual há mais dúvidas do que certezas.

A má vontade com o debate era tão grande que a audiência ocorreu por meios virtuais, no ano passado, quando a pandemia já havia arrefecido. O Instituto Tecnologia e Sociedade, que participou da discussão, disse à época que as sugestões apresentadas pelos pesquisadores foram ignoradas. Políticos adoram assinar contrato e são reticentes ou mudos quando são questionados.

O projeto de São Paulo era tão primário que havia barbaridades na 1ª versão apresentada ao público. Dizia-se que o reconhecimento facial serviria para combater “vadiagem”.

O problema é que vadiagem não é crime. É uma contravenção prevista num decreto-lei de 1941, criado na ditadura de Getúlio Vargas. São Paulo queria usá-la contra moradores de rua, mas pegou tão mal querer aplicar tecnologia contra quem não tem nada que a prefeitura tirou a vadiagem do projeto. A Justiça chegou a suspender a licitação por duas vezes, uma das quais por conta do risco de racismo.

Não é por falta de pesquisas que se deixou de discutir os graves riscos do reconhecimento facial. Há centenas delas, mas vou destacar duas para mostrar o risco que São Paulo está correndo, inclusive de ter de pagar indenizações milionárias pelos erros que pode cometer.

Em 2018, a União Americana pelos Direitos Civis, uma ONG que tem 103 anos, usou o sistema de reconhecimento facial da Amazon, o Rekognition, para testar a acuidade do sistema. Pegou fotos de todos os congressistas em Washington e jogou-as num sistema que tinha 25.000 imagens de pessoas presas nos EUA. O sistema da Amazon “reconheceu” 28 membros do Congresso como sendo aqueles criminosos que estavam presos.

Há um viés racista no reconhecimento facial da Amazon, segundo o levantamento da ONG: em 40% dos casos de falso reconhecimento as pessoas eram negras – apesar de o Congresso ter só 20% de políticos negros.

Uma pesquisa feita no MIT em 2018 mostrou um novo fenômeno na internet: o racismo algoritmo. Joy Buolamwini, artista e pesquisadora do MIT Midia Lab, fez um levantamento revelador junto com um pesquisador da Microsoft, Timnit Gebru: eles experimentaram o reconhecimento de rostos negros nos sistemas da Amazon, Google, IBM e da própria Microsoft.

O resultado foi que as mulheres negras são as que têm mais falsos positivos (como eles chamam os erros): elas não são reconhecidas em 34,7% dos casos. Já entre os homens brancos, os casos de falso positivo são 0,8%.

O motivo do racismo algoritmo é simples: os sistemas de aprendizado de máquina são feitos usando fotos de brancos, principalmente de homens. A revelação obrigou as empresas a mudarem o sistema de aprendizado de máquina.

Joy Buolamwini virou uma estrela após a sua descoberta. Ela é protagonista do documentário da Netflix chamado “Preconceito Codificado” (Coded Bias, 2022).

Se o prefeito de São Paulo quiser economizar em indenizações futuras, devia dar uma espiada no doc.

Anexo IV – Ocorrência 4

Notícias e Sociedade

Inocentado anteriormente, homem é acusado de roubo de novo por reconhecimento facial

Publicado:10 de abril de 202320:06 Atualizado: 20:58

Mariane Del Rei

Em 2020, o educador Danillo Félix de Oliveira foi preso, acusado de roubo injustamente, e depois de dois meses foi solto, pois provou sua inocência. Agora, o pesadelo se repete. Danilo se tornou réu novamente pelo método de reconhecimento facial, no estado do Rio de Janeiro.

Mesmo depois da própria vítima do assalto reconhecer o "engano", a foto de Danillo continuou no banco de dados de suspeitos da 76ª DP (Niterói). Desta vez, Danilo responde a acusação de roubo em liberdade, ao contrário do que aconteceu em 2020, quando foi detido no meio da rua e ficou encarcerado por dois meses, em três presídios diferentes.

Danillo questiona reconhecimento facial, método que o tornou réu pela segunda vez. Foto: Reprodução/TV Globo

O Instituto de Defesa da População Negra assumiu a defesa do caso. Danillo questionou o método que o tornou réu pela segunda vez. "A dinâmica é a mesma. Parece que no mesmo dia essa pessoa [o verdadeiro assaltante] cometeu outros assaltos. Os processos são os mesmos, a juíza é a mesma, a vara é a mesma. Eu já fui julgado. Para que eu vou passar por outra audiência novamente?", disse o educador à imprensa.

"Sou um jovem negro de classe baixa, morador de comunidade. Não tem prova nenhuma para me prenderem, e me prenderam. Isso não é racismo? Óbvio que é. Não de uma pessoa específica, mas do Estado, desse Brasil", concluiu.

Leia também: Prefeitura de São Paulo deve relançar programa de reconhecimento facial

Danillo foi denunciado com base no registro de ocorrência da delegacia e as vítimas vão ser ouvidas em audiência nesta semana, segundo o Ministério Público do Rio de Janeiro (MPRJ). O Tribunal de Justiça afirmou que Danillo é réu e responde por roubo majorado.

Racismo algoritmo

Segundo a Rede de Observatório da Segurança, 90,5% das prisões feitas através do reconhecimento facial foram de pessoas negras. Alguns nunca tiveram passagem pela polícia e não sabiam como passaram a integrar o banco de dados de criminosos. Os outros 9,5% são ocupados por pessoas brancas

Já um levantamento realizado pela Defensoria Pública do Rio de Janeiro mostrou que 80% dos réus absolvidos por erros na identificação feita por reconhecimento fotográfico passaram, em média, um ano e dois meses presos antes do julgamento.

O estudo analisou 242 processos julgados no Tribunal de Justiça do estado entre os meses de janeiro e junho de 2021. Em um dos casos, o acusado passou quase seis anos encarcerado preventivamente até a absolvição.

Os erros na identificação fotográfica foram atribuídos à utilização de um álbum de suspeitos que podem conter fotos de indivíduos existentes na delegacia ou obtidos nas redes sociais. Na maior parte dos casos, as pessoas que constam no álbum não respondem a nenhum crime perante a Justiça.

A pesquisa ainda apontou o perfil dos acusados com base no reconhecimento fotográfico. Em geral, são homens e negros. Este é o mesmo perfil entre os réus julgados: 95,9% são homens e 63,74%, negros, somando-se pretos e pardos conforme a definição do IBGE.

Anexo V – Ocorrência 5

JORNAL DO BRASIL
desde 1891
BRASIL

Feia x bonita: como o racismo algorítmico impacta imagem de mulheres negras na internet

Estudos apontam as consequências e possíveis soluções para moderar conteúdos discriminatórios reproduzidos no ambiente virtual

Por JORNAL DO BRASIL com Alma Preta Jornalismo redacao@jb.com.br. Publicado em 26/07/2023 às 05:52/Alterado em 26/07/2023 às 08:22

A pesquisadora Emilly Lima Arquivo Pessoal

"Cabelo ruim", "cabelo bom", "pele feia" e "pele bonita". Você já prestou atenção nas imagens que aparecem na internet quando se busca por essas palavras? É comum encontrar fotos de mulheres negras quando as palavras estão associadas a algo negativo, enquanto as expressões positivas são direcionadas majoritariamente a imagens de mulheres brancas. Em uma rápida pesquisa no Google, um dos maiores buscadores do mundo, ainda é possível visualizar atribuições negativas a mulheres negras. Mas por que isso acontece? Pesquisadores do campo da tecnologia chamam de "racismo algorítmico" o conceito utilizado para nomear a reprodução do racismo estrutural dentro do ambiente digital, conforme detalha o especialista Tarcízio Silva, um dos principais estudiosos do termo no Brasil.

"O racismo algoritmo é o modo pelo qual o racismo estrutural é atualizado e reproduzido por tecnologias digitais algorítmicas, que geralmente são chamadas de inteligência artificial — mas eu não gosto de usar esse termo — ao incorporar decisões que podem ser decisões discriminatórias geralmente para aumento de lucro e atividade das empresas de uma forma irresponsável e, em alguns casos, de forma intencional devido à supremacia branca e ao racismo", destaca Silva.

No Brasil, um dos casos emblemáticos sobre os impactos da reprodução do racismo e sexismo nos buscadores aconteceu em 2019, quando a empresária baiana e relações públicas Cáren Cruz pesquisou na internet "mulher negra dando aula" e se deparou com imagens que associavam mulheres negras a conteúdos pornográficos. Ela preparava uma apresentação corporativa para uma empresa e tinha feito a pesquisa porque só encontrava imagens de mulheres brancas em posições de ensino. O resultado das imagens explícitas já foi retirado do buscador pela plataforma.

Com a repercussão, na época o Google informou ao site Bahia Notícias que também tinha sido surpreendido e reconheceu que as imagens não deveriam ficar explícitas. "Quando as pessoas usam a busca, queremos oferecer resultados relevantes para os termos usados nas pesquisas e não temos a intenção de mostrar resultados explícitos para os usuários, a não ser que estejam buscando isso. Claramente, o conjunto de resultados para o termo mencionado não está à altura desse princípio e pedimos desculpas àqueles que se sentiram impactados ou ofendidos", escreveu a empresa em nota enviada ao site.

Apesar dos buscadores usarem o argumento de que os resultados são frutos dos conteúdos com maior relevância ou por palavras-chave, a professora e pesquisadora estadunidense e uma das expoentes do conceito do "racismo algorítmico", Safiya Noble, questiona a suposta "neutralidade" dos buscadores na categorização das pesquisas.

No livro "Algoritmos da opressão: como os mecanismos de busca reforçam o racismo", lançado em 2021, Noble chama atenção para a reprodução das desigualdades estruturais no ambiente virtual. "Desigualdades estruturais da sociedade estão sendo reproduzidas na internet e a luta por um espaço cibernético sem raça, gênero e classes pode apenas 'perpetuar e reforçar os atuais sistemas de dominação'", cita a especialista, que também atuou na área de marketing por mais de dez anos.

A reportagem entrou em contato com a equipe do Google no Brasil e questionou sobre como funciona o algoritmo de buscas da plataforma e ações/estudos que a empresa adota para evitar e corrigir a reprodução de resultados nocivos a grupos historicamente discriminados. Em nota, o Google informou que pelo fato de os sistemas serem organizados com base na "internet aberta", a plataforma pode refletir preconceitos já rotulados na internet.

"Compartilhamos a profunda preocupação em torno disso e estamos trabalhando ativamente em soluções escaláveis para esses tipos de problemas, tanto que já tivemos melhorias significativas nos últimos anos. Como a internet está em constante mudança, esse é um desafio contínuo e continuaremos trabalhando para superá-lo como parte do nosso compromisso de criar produtos úteis e inclusivos para todos os usuários", completa a nota.

O Google também citou que em maio do ano passado anunciou o lançamento da "Escala Monk Skin Tone (MST)", projetada para incluir mais o espectro de tons de pele na pesquisa de imagens da plataforma. A ferramenta foi produzida com base na pesquisa do professor e sociólogo de Harvard, Dr. Ellis Monk, que estuda como o tom de pele e o colorismo afetam a vida das pessoas há mais de 10 anos.

Mulheres negras são as mais hostilizadas no ambiente virtual

Apesar de não existirem dados específicos sobre racismo algorítmico no Brasil, pesquisas indicam que o ambiente virtual é mais hostil para mulheres negras. A tese de doutorado do pesquisador e PHD em Sociologia, Luiz Valério Trindade, aponta que as mulheres negras correspondem a 81% das vítimas de discurso discriminatório nas redes sociais. O perfil da maioria (65%) dos internautas que disseminam intolerância racial é composto por homens, na faixa de 20 a 25 anos.

No contexto das últimas eleições municipais, em 2020, o Instituto Marielle Franco realizou uma pesquisa inédita sobre violência política e os resultados mostram que as candidatas negras foram as

que mais sofreram com a violência virtual, relatada por 78% das entrevistadas. Em seguida, estão a violência moral e psicológica (62%), violência institucional (55%) e violência racial (44%).

Segundo o pesquisador Tarcízio Silva, a falta de transparência das plataformas digitais é um dos fatores que dificultam a elaboração de estratégias e o mapeamento do racismo algorítmico no Brasil.

"As plataformas digitais, quando a gente está falando da internet, não oferecem transparência sobre esses tipos de informações, basicamente sobre quase nenhum tipo de formação relevante para a sociedade e sobre danos possíveis, seja danos sobre discriminação, moderação de conteúdo inadequada, desinformação etc. E aí o que está em jogo hoje na regulação de plataformas, por exemplo, envolve forçar as plataformas a oferecerem dados que estão relacionados a isso. No Brasil eu diria que não há dados quantitativos sobre racismo algorítmico", ressalta Silva.

Quando se pensa no conceito do “racismo algorítmico”, os bancos de imagens também não estão isentos de reproduzir desigualdades estruturais. É o que indica uma pesquisa realizada pela psicóloga Emilly Lima, integrante do grupo de pesquisa em Percepção Visual da Universidade de Brasília (UnB).

Ainda em produção, o estudo analisou mais de 3 mil fotos em bancos de imagens digitais gratuitos para investigar e analisar a representação racial e socioeconômica presente na plataforma ao pesquisar por quatro palavras-chave: riqueza e dinheiro, para indicar alto status social; e pobreza e miséria, para indicar baixo status social. Os resultados parciais apontam que a maioria dos representados na pesquisa de baixo status social foram pessoas negras.

"Quando se busca por riqueza e pobreza, tem uma diferença na forma como as pessoas são representadas. Para ‘riqueza’ a gente identificou mais imagens como barras de ouro, uma casa, um carro, mas quando se olha para ‘pobreza’ geralmente são pessoas em situação de pobreza e, em sua maioria, pessoas negras nessa situação", explica Emilly Lima.

A pesquisadora ressalta que a maioria das fotos dos bancos de imagens são da Europa e da América do Norte, por isso existe uma maior quantidade de pessoas brancas em termos de proporção do que de pessoas negras, e quando as pessoas negras aparecem, geralmente, são em situações de baixo status socioeconômico.

Emily avalia ser necessário combater o discurso de "neutralidade" dos algoritmos. "Acho que o discurso da neutralidade cai quando a gente vê que a internet é formada por pessoas, que existem pessoas por trás da internet. Quem é que programa as máquinas? Quem é que está por trás da programação dessa máquina? A gente está usando banco de imagens diversos que representa toda a população do mundo ou a gente só está massificando essa cultura de acordo com o padrão branco e eurocêntrico?", questiona a pesquisadora.

Possíveis soluções e desafios

O debate sobre o racismo algorítmico também perpassa pelo debate sobre a regulação das redes sociais no Brasil, segundo especialistas ouvidos pela Alma Preta Jornalismo. Atualmente, o Marco Civil da Internet, através da lei 12.965/14, regula as redes sociais no país, no entanto, não

estabelece a responsabilização das plataformas digitais, que se posicionam apenas como "intermediárias" para os usuários.

Proposta durante a campanha do então candidato Luís Inácio Lula da Silva (PT) e atual presidente, a discussão sobre a ampliação da regulação tem como foco central combater a desinformação, discursos de ódio e disseminação de informações que ameaçam a democracia. Entretanto, ainda há desafios a serem enfrentados nesse campo como, por exemplo, a falta de participação de pessoas negras, as atribuições e responsabilidades das plataformas e quem irá monitorar e regular.

De acordo com Tarcízio, um dos debates principais deve ser estabelecer quais são os riscos possíveis nesse ambiente, quais as obrigações das plataformas e os riscos considerados inaceitáveis que devem ser analisados na regulação das plataformas.

"O debate que alguns movimentos, sobretudo o movimento negro, tem tentado fazer é estabelecer que, por exemplo, o reconhecimento facial em espaço público é um risco inaceitável e a partir daí, se ela é uma tecnologia de risco alto, não deve ser desenvolvida. Só que as perspectivas liberais, inclusive na sociedade civil, dificultam essa posição e em termos de risco alto as empresas, estados ou governos teriam obrigações de transparência, responsabilidade e de reparação", comenta.

Segundo a coordenadora de pesquisa e pesquisadora do Instituto de Referência em Internet e Sociedade (IRIS), Fernanda Rodrigues, apesar das ferramentas para o combate ao “racismo algorítmico”, ainda há a necessidade em ter instrumentos normativos importantes para reconhecer esse conceito para o cumprimento da lei e mecanismos de fiscalização.

"Em relação à regulação da inteligência artificial, por exemplo, a gente poderia ter mecanismos de fiscalização, como a avaliação do impacto algorítmico que considere as questões de raça, gênero, dentre outras, para analisar se aquela ferramenta da inteligência artificial vai ter impacto sobre determinados grupos sociais, como mulheres e pessoas negras", destaca Fernanda, também doutoranda em Direito pela Universidade Federal de Minas Gerais (UFMG) na área de Direito, Tecnociências e Interdisciplinaridade.

Para o avanço do debate sobre a regulamentação no país, a especialista avalia que apesar da contribuição de discussões já estabelecidas no contexto europeu, a exemplo da União Europeia, é importante destacar narrativas que estão para além do eixo dominante de produção e epistemologia, como as regulamentações discutidas no sul global.

"A Declaração de Windhoek sobre a inteligência artificial em países do sul da África fala justamente sobre perspectivas relacionadas à necessidade de uma decolonização da tecnologia — do sistema da inteligência artificial especificamente — e também numa decolonização da própria educação em todos os níveis relacionados a essa tecnologia para que gente possa realmente pensar nesse tipo de ferramenta a partir do nosso contexto e das nossas demandas", conclui.

RESEARCH ARTICLE

ECONOMICS

Dissecting racial bias in an algorithm used to manage the health of populations

Ziad Obermeyer[1,2]*, Brian Powers[3], Christine Vogeli[4], Sendhil Mullainathan[5]*†

Health systems rely on commercial prediction algorithms to identify and help patients with complex health needs. We show that a widely used algorithm, typical of this industry-wide approach and affecting millions of patients, exhibits significant racial bias: At a given risk score, Black patients are considerably sicker than White patients, as evidenced by signs of uncontrolled illnesses. Remedying this disparity would increase the percentage of Black patients receiving additional help from 17.7 to 46.5%. The bias arises because the algorithm predicts health care costs rather than illness, but unequal access to care means that we spend less money caring for Black patients than for White patients. Thus, despite health care cost appearing to be an effective proxy for health by some measures of predictive accuracy, large racial biases arise. We suggest that the choice of convenient, seemingly effective proxies for ground truth can be an important source of algorithmic bias in many contexts.

There is growing concern that algorithms may reproduce racial and gender disparities via the people building them or through the data used to train them (*1*–*3*). Empirical work is increasingly lending support to these concerns. For example, job search ads for highly paid positions are less likely to be presented to women (*4*), searches for distinctively Black-sounding names are more likely to trigger ads for arrest records (*5*), and image searches for professions such as CEO produce fewer images of women (*6*). Facial recognition systems increasingly used in law enforcement perform worse on recognizing faces of women and Black individuals (*7*, *8*), and natural language processing algorithms encode language in gendered ways (*9*).

Empirical investigations of algorithmic bias, though, have been hindered by a key constraint: Algorithms deployed on large scales are typically proprietary, making it difficult for independent researchers to dissect them. Instead, researchers must work "from the outside," often with great ingenuity, and resort to clever workarounds such as audit studies. Such efforts can document disparities, but understanding how and why they arise—much less figuring out what to do about them—is difficult without greater access to the algorithms themselves. Our understanding of a mechanism therefore typically relies on theory or exercises with researcher-created algorithms (*10*–*13*). Without an algorithm's training data, objective function, and prediction methodology, we can only guess as to the actual mechanisms for the important algorithmic disparities that arise.

In this study, we exploit a rich dataset that provides insight into a live, scaled algorithm deployed nationwide today. It is one of the largest and most typical examples of a class of commercial risk-prediction tools that, by industry estimates, are applied to roughly 200 million people in the United States each year. Large health systems and payers rely on this algorithm to target patients for "high-risk care management" programs. These programs seek to improve the care of patients with complex health needs by providing additional resources, including greater attention from trained providers, to help ensure that care is well coordinated. Most health systems use these programs as the cornerstone of population health management efforts, and they are widely considered effective at improving outcomes and satisfaction while reducing costs (*14*–*17*). Because the programs are themselves expensive—with costs going toward teams of dedicated nurses, extra primary care appointment slots, and other scarce resources—health systems rely extensively on algorithms to identify patients who will benefit the most (*18*, *19*).

Identifying patients who will derive the greatest benefit from these programs is a challenging causal inference problem that requires estimation of individual treatment effects. To solve this problem, health systems make a key assumption: Those with the greatest care needs will benefit the most from the program. Under this assumption, the targeting problem becomes a pure prediction policy problem (*20*). Developers then build algorithms that rely on past data to build a predictor of future health care needs.

Our dataset describes one such typical algorithm. It contains both the algorithm's predictions as well as the data needed to understand its inner workings: that is, the underlying ingredients used to form the algorithm (data, objective function, etc.) and links to a rich set of outcome data. Because we have the inputs, outputs, and eventual outcomes, our data allow us a rare opportunity to quantify racial disparities in algorithms and isolate the mechanisms by which they arise. It should be emphasized that this algorithm is not unique. Rather, it is emblematic of a generalized approach to risk prediction in the health sector, widely adopted by a range of for- and non-profit medical centers and governmental agencies (*21*).

Our analysis has implications beyond what we learn about this particular algorithm. First, the specific problem solved by this algorithm has analogies in many other sectors: The predicted risk of some future outcome (in our case, health care needs) is widely used to target policy interventions under the assumption that the treatment effect is monotonic in that risk, and the methods used to build the algorithm are standard. Mechanisms of bias uncovered in this study likely operate elsewhere. Second, even beyond our particular finding, we hope that this exercise illustrates the importance, and the large opportunity, of studying algorithmic bias in health care, not just as a model system but also in its own right. By any standard—e.g., number of lives affected, life-and-death consequences of the decision—health is one of the most important and widespread social sectors in which algorithms are already used at scale today, unbeknownst to many.

Data and analytic strategy

Working with a large academic hospital, we identified all primary care patients enrolled in risk-based contracts from 2013 to 2015. Our primary interest was in studying differences between White and Black patients. We formed race categories by using hospital records, which are based on patient self-reporting. Any patient who identified as Black was considered to be Black for the purpose of this analysis. Of the remaining patients, those who self-identified as races other than White (e.g., Hispanic) were so considered (data on these patients are presented in table S1 and fig. S1 in the supplementary materials). We considered all remaining patients to be White. This approach allowed us to study one particular racial difference of social and historical interest between patients who self-identified as Black and patients who self-identified as White without another race or ethnicity; it has the disadvantage of not allowing for the study of intersectional racial

[1]School of Public Health, University of California, Berkeley, Berkeley, CA, USA. [2]Department of Emergency Medicine, Brigham and Women's Hospital, Boston, MA, USA. [3]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [4]Mongan Institute Health Policy Center, Massachusetts General Hospital, Boston, MA, USA. [5]Booth School of Business, University of Chicago, Chicago, IL, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: sendhil.mullainathan@chicagobooth.edu

and ethnic identities. Our main sample thus consisted of (i) 6079 patients who self-identified as Black and (ii) 43,539 patients who self-identified as White without another race or ethnicity, whom we observed over 11,929 and 88,080 patient-years, respectively (1 patient-year represents data collected for an individual patient in a calendar year). The sample was 71.2% enrolled in commercial insurance and 28.8% in Medicare; on average, 50.9 years old; and 63% female (Table 1).

For these patients, we obtained algorithmic risk scores generated for each patient-year. In the health system we studied, risk scores are generated for each patient during the enrollment period for the system's care management program. Patients above the 97th percentile are automatically identified for enrollment in the program. Those above the 55th percentile are referred to their primary care physician, who is provided with contextual data about the patients and asked to consider whether they would benefit from program enrollment.

Many existing metrics of algorithmic bias may apply to this scenario. Some definitions focus on calibration [i.e., whether the realized value of some variable of interest Y matches the risk score R (*2*, *22*, *23*)]; others on statistical parity of some decision D influenced by the algorithm (*10*); and still others on balance of average predictions, conditional on the realized outcome (*22*). Given this multiplicity and the growing recognition that not all conditions can be simultaneously satisfied (*3*, *10*, *22*), we focus on metrics most relevant to the real-world use of the algorithm, which are related to calibration bias [formally, comparing Blacks B and Whites W, $E[Y|R, W] = E[Y|R, B]$ indicates the absence of bias (here, E is the expectation operator)]. The algorithm's stated goal is to predict complex health needs for the purpose of targeting an intervention that manages those needs. Thus, we compare the algorithmic risk score for patient i in year t ($R_{i,t}$), formed on the basis of claims data $X_{i,(t-1)}$ from the prior year, to data on patients' realized health $H_{i,t}$, assessing how well the algorithmic risk score is calibrated across race for health outcomes $H_{i,t}$. We also ask how well the algorithm is calibrated for costs $C_{i,t}$.

To measure H, we link predictions to a wide range of outcomes in electronic health record data, including all diagnoses (in the form of International Classification of Diseases codes) as well as key quantitative laboratory studies and vital signs capturing the severity of chronic illnesses. To measure C, we link predictions to insurance claims data on utilization, including outpatient and emergency visits, hospitalizations, and health care costs. These data, and the rationale for the specific measures of H used in this study, are described in more detail in the supplementary materials.

Health disparities conditional on risk score

We begin by calculating an overall measure of health status, the number of active chronic conditions [or "comorbidity score," a metric used extensively in medical research (*24*) to provide a comprehensive view of a patient's health (*25*)] by race, conditional on algorithmic risk score. Fig. 1A shows that, at the same level of algorithm-predicted risk, Blacks have significantly more illness burden than Whites. We can quantify these differences by choosing one point on the x axis that corresponds to a very-high-risk group (e.g., patients at the 97th percentile of risk score, at which patients are auto-identified for program enrollment), where Blacks have 26.3% more chronic illnesses than Whites (4.8 versus 3.8 distinct conditions; $P < 0.001$).

What do these prediction differences mean for patients? Algorithm scores are a key input to decisions about future enrollment in a care coordination program. So as we might expect, with less-healthy Blacks scored at similar risk scores to more-healthy Whites, we find evidence

Table 1. Descriptive statistics on our sample, by race. BP, blood pressure; LDL, low-density lipoprotein.

	White	Black
n (patient-years)	88,080	11,929
n (patients)	43,539	6079
Demographics		
Age	51.3	48.6
Female (%)	62	69
Care management program		
Algorithm score (percentile)	50	52
Race composition of program (%)	81.8	18.2
Care utilization		
Actual cost	$7540	$8442
Hospitalizations	0.09	0.13
Hospital days	0.50	0.78
Emergency visits	0.19	0.35
Outpatient visits	4.94	4.31
Mean biomarker values		
HbA1c (%)	5.9	6.4
Systolic BP (mmHg)	126.6	130.3
Diastolic BP (mmHg)	75.5	75.7
Creatinine (mg/dl)	0.89	0.98
Hematocrit (%)	40.7	37.8
LDL (mg/dl)	103.4	103.0
Active chronic illnesses (comorbidities)		
Total number of active illnesses	1.20	1.90
Hypertension	0.29	0.44
Diabetes, uncomplicated	0.08	0.22
Arrythmia	0.09	0.08
Hypothyroid	0.09	0.05
Obesity	0.07	0.18
Pulmonary disease	0.07	0.11
Cancer	0.07	0.06
Depression	0.06	0.08
Anemia	0.05	0.10
Arthritis	0.04	0.04
Renal failure	0.03	0.07
Electrolyte disorder	0.03	0.05
Heart failure	0.03	0.05
Psychosis	0.03	0.05
Valvular disease	0.03	0.02
Stroke	0.02	0.03
Peripheral vascular disease	0.02	0.02
Diabetes, complicated	0.02	0.07
Heart attack	0.01	0.02
Liver disease	0.01	0.02

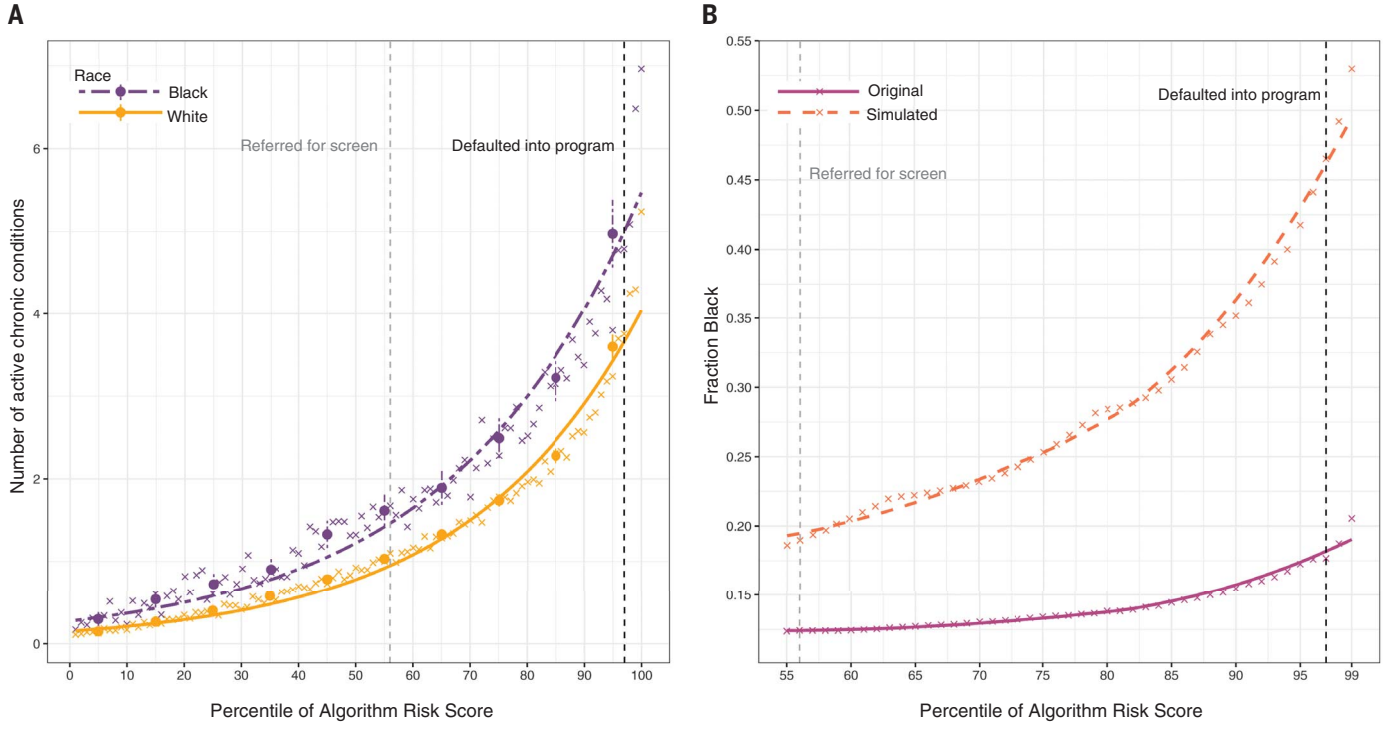


Fig. 1. Number of chronic illnesses versus algorithm-predicted risk, by race. (**A**) Mean number of chronic conditions by race, plotted against algorithm risk score. (**B**) Fraction of Black patients at or above a given risk score for the original algorithm ("original") and for a simulated scenario that removes algorithmic bias ("simulated": at each threshold of risk, defined at a given percentile on the *x* axis, healthier Whites above the threshold are replaced with less healthy Blacks below the threshold, until the marginal patient is equally healthy). The × symbols show risk percentiles by race; circles show risk deciles with 95% confidence intervals clustered by patient. The dashed vertical lines show the auto-identification threshold (the black line, which denotes the 97th percentile) and the screening threshold (the gray line, which denotes the 55th percentile).

of substantial disparities in program screening. We quantify this by simulating a counterfactual world with no gap in health conditional on risk. Specifically, at some risk threshold α, we identify the supramarginal White patient (i) with $R_i > \alpha$ and compare this patient's health to that of the inframarginal Black patient (j) with $R_j < \alpha$. If $H_i > H_j$, as measured by number of chronic medical conditions, we replace the (healthier, but supramarginal) White patient with the (sicker, but inframarginal) Black patient. We repeat this procedure until $H_i = H_j$, to simulate an algorithm with no predictive gap between Blacks and Whites. Fig. 1B shows the results: At all risk thresholds α above the 50th percentile, this procedure would increase the fraction of Black patients. For example, at α = 97th percentile, among those auto-identified for the program, the fraction of Black patients would rise from 17.7 to 46.5%.

We then turn to a more multidimensional picture of the complexity and severity of patients' health status, as measured by biomarkers that index the severity of the most common chronic illnesses in our sample (as shown in Table 1). This allows us to identify patients who might derive a great deal of benefit from care management programs—e.g., patients with severe diabetes who are at risk of catastrophic complications if they do not lower their blood sugar (*18*, *26*). (The materials and methods section describes several experiments to rule out a large effect of the program on these health measures in year t; had there been such an effect, we could not easily use the measures to assess the accuracy of the algorithm's predictions on health, because the program is allocated as a function of algorithm score.) Across all of these important markers of health needs—severity of diabetes, high blood pressure, renal failure, cholesterol, and anemia—we find that Blacks are substantially less healthy than Whites at any level of algorithm predictions, as shown in Fig. 2. Blacks have more-severe hypertension, diabetes, renal failure, and anemia, and higher cholesterol. The magnitudes of these differences are large: For example, differences in severity of hypertension (systolic pressure: 5.7 mmHg) and diabetes [glycated hemoglobin (HbA1c): 0.6%] imply differences in all-cause mortality of 7.6% (*27*) and 30% (*28*), respectively, calculated using data from clinical trials and longitudinal studies.

Mechanism of bias

An unusual aspect of our dataset is that we observe the algorithm's inputs and outputs as well as its objective function, providing us a unique window into the mechanisms by which bias arises. In our setting, the algorithm takes in a large set of raw insurance claims data $X_{i,t-1}$ (features) over the year $t-1$: demographics (e.g., age, sex), insurance type, diagnosis and procedure codes, medications, and detailed costs. Notably, the algorithm specifically excludes race.

The algorithm uses these data to predict $Y_{i,t}$ (i.e., the label). In this instance, the algorithm takes total medical expenditures (for simplicity, we denote "costs" C_t) in year t as the label. Thus, the algorithm's prediction on health needs is, in fact, a prediction on health costs.

As a first check on this potential mechanism of bias, we calculate the distribution of realized costs C versus predicted costs R. By this metric, one could call the algorithm unbiased. Fig. 3A shows that, at every level of algorithm-predicted risk, Blacks and Whites have (roughly) the same costs the following year. In other words, the algorithm's predictions are well calibrated across races. For example, at the median risk score, Black patients had costs of $5147 versus $4995 for Whites (U.S. dollars); in the top 5% of algorithm-predicted risk, costs were $35,541 for Blacks versus $34,059 for Whites.

Because these programs are used to target patients with high costs, these results are largely inconsistent with algorithmic bias, as measured by calibration: Conditional on risk score, predictions do not favor Whites or Blacks anywhere in the risk distribution.

To summarize, we find substantial disparities in health conditional on risk but little disparity in costs. On the one hand, this is surprising: Health care costs and health needs are highly correlated, as sicker patients need and receive more care, on average. On the other hand, there are many opportunities for a wedge to creep in between needing health care and receiving health care—and crucially, we find that wedge to be correlated with race, as shown in Fig. 3B. At a given level of health (again measured by number of chronic illnesses), Blacks generate lower costs than Whites—on average, $1801 less per year, holding constant the number of chronic illnesses (or $1144 less, if we instead hold constant the specific individual illnesses that contribute to the sum). Table S2 also shows that Black patients generate very different kinds of costs: for example, fewer inpatient surgical and outpatient specialist costs, and more costs related to emergency visits and dialysis. These results suggest that the driving force behind the bias we detect is that Black patients generate lesser medical expenses, conditional on health, even when we account for specific comorbidities. As a result, accurate prediction of costs necessarily means being racially biased on health.

How might these disparities in cost arise? The literature broadly suggests two main potential channels. First, poor patients face substantial barriers to accessing health care, even when enrolled in insurance plans. Although the population we study is entirely insured, there are many other mechanisms by which poverty can lead to disparities in use of health care: geography and differential access to transportation, competing demands from jobs or child care, or knowledge of reasons to seek care (*29*–*31*). To the extent that race and socioeconomic status are correlated, these factors will differentially affect Black patients. Second, race could affect costs directly via several channels: direct ("taste-based") discrimination, changes to the doctor–patient relationship, or others. A recent trial randomly assigned Black patients to a Black or White primary care provider and found significantly higher uptake of recommended preventive care when the provider was Black (*32*). This is perhaps the most rigorous demonstration of this effect, and it fits with a larger literature on potential mechanisms by which race can affect health care directly. For example, it has long been documented that Black patients have reduced trust in the health care system (*33*), a fact that some studies trace to the revelations of the Tuskegee study and other adverse experiences (*34*). A substantial literature in psychology has documented physicians' differential perceptions of Black patients, in terms of intelligence, affiliation (*35*), or pain tolerance (*36*). Thus, whether it is communication, trust, or bias, something about the interactions of Black patients with the health care system itself leads to reduced use of health care. The collective effect of these many channels is to lower health spending substantially for Black patients, conditional on need—a finding that has been appreciated for at least two decades (*37*).

Problem formulation

Our findings highlight the importance of the choice of the label on which the algorithm is trained. On the one hand, the algorithm manufacturer's choice to predict future costs is reasonable: The program's goal, at least in part, is

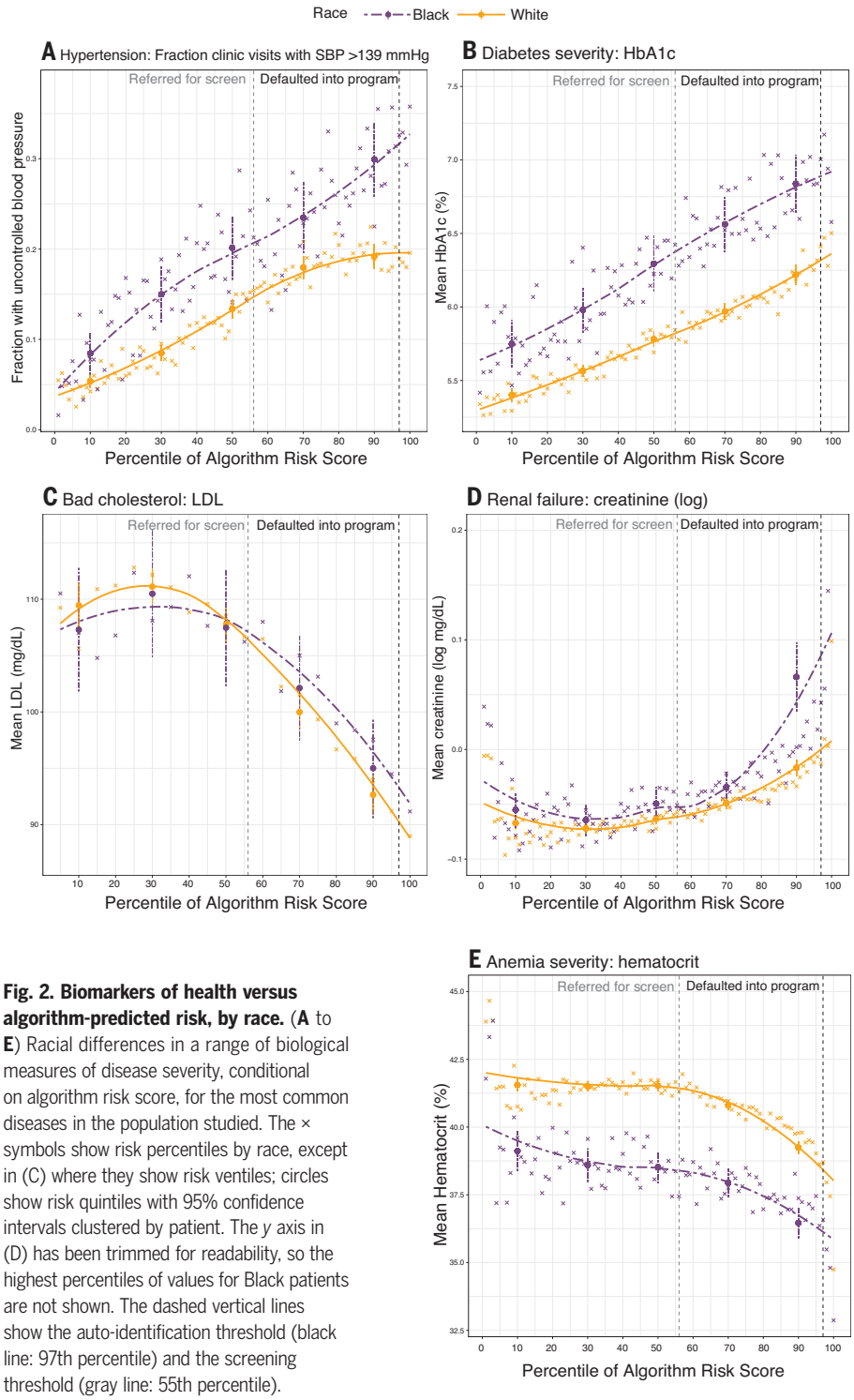


Fig. 2. Biomarkers of health versus algorithm-predicted risk, by race. (**A** to **E**) Racial differences in a range of biological measures of disease severity, conditional on algorithm risk score, for the most common diseases in the population studied. The × symbols show risk percentiles by race, except in (C) where they show risk ventiles; circles show risk quintiles with 95% confidence intervals clustered by patient. The *y* axis in (D) has been trimmed for readability, so the highest percentiles of values for Black patients are not shown. The dashed vertical lines show the auto-identification threshold (black line: 97th percentile) and the screening threshold (gray line: 55th percentile).

to reduce costs, and it stands to reason that patients with the greatest future costs could have the greatest benefit from the program. As noted in the supplementary materials, the manufacturer is not alone. Although the details of individual algorithms vary, the cost label reflects the industry-wide approach. For example, the Society of Actuaries's comprehensive evaluation of the 10 most widely used algorithms, including the particular algorithm we study, used cost prediction as its accuracy metric (*21*). As noted in the report, the enthusiasm for cost prediction is not restricted to industry: Similar algorithms are developed and used by non-profit hospitals, academic groups, and governmental agencies, and are often described in academic literature on targeting population health interventions (*18*, *19*).

On the other hand, future cost is by no means the only reasonable choice. For example, the evidence on care management programs shows that they do not operate to reduce costs globally. Rather, these programs primarily work to prevent acute health decompensations that lead to catastrophic health care utilization (indeed, they actually work to increase other categories of costs, such as primary care and home health assistance; see table S2). Thus avoidable future costs, i.e., those related to emergency visits and hospitalizations, could be a useful label to predict. Alternatively, rather than predicting costs at all, we could simply predict a measure of health; e.g., the number of active chronic health conditions. Because the program ultimately operates to improve the management of these conditions, patients with the most encounters related to them could also be a promising group on which to deploy preventative interventions.

The dilemma of which label to choose relates to a growing literature on "problem formulation" in data science: the task of turning an often amorphous concept we wish to predict into a concrete variable that can be predicted in a given dataset (*38*). Problems in health seem particularly challenging: Health is, by nature, holistic and multidimensional, and there is no single, precise way to measure it. Health care costs, though well measured and readily available in insurance claims data, are also the result of a complex aggregation process with a number of distortions due to structural inequality, incentives, and inefficiency. So although the choice of label is perhaps the single most important decision made in the development of a prediction algorithm, in our setting and in many others, there is often a confusingly large array of different options, each with its own profile of costs and benefits.

Experiments on label choice

Through a series of experiments with our dataset, we can gain some insight into how label choice affects both predictive performance and racial bias. We develop three new predictive algorithms, all trained in the same way, to predict the following outcomes: total cost in year t (this tailors cost predictions to our own dataset rather than the national training set), avoidable cost in year t (due to emergency visits and hospitalizations), and health in year t (measured by the number of chronic conditions that flare up in that year). We train all models in a random ⅔ training set and show all results only from the ⅓ holdout set. Furthermore, as with the original algorithm, we exclude race from the feature set (more details are in the materials and methods).

Table 2 shows the results of these experiments. The first finding is that all algorithms perform reasonably well for predicting not only the outcome on which they were trained but also the other outcomes: The concentration of realized outcomes in those at or above the 97th percentile is notably similar for all algorithms across all outcomes. The largest difference in performance across algorithms is seen for cost prediction: Of all costs in the holdout set, the fraction generated by those at or above the 97th percentile is 16.5% for the cost predictor versus 12.1% for the predictor

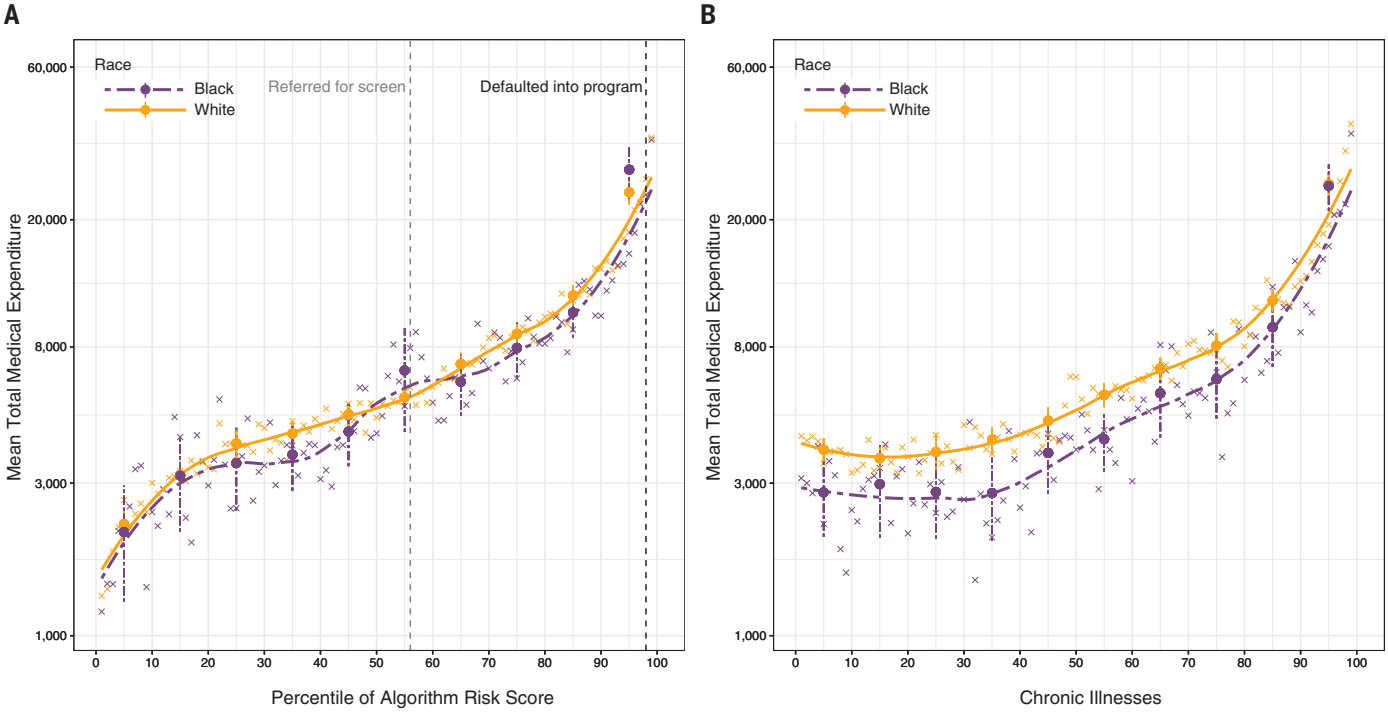


Fig. 3. Costs versus algorithm-predicted risk, and costs versus health, by race. (**A**) Total medical expenditures by race, conditional on algorithm risk score. The dashed vertical lines show the auto-identification threshold (black line: 97th percentile) and the screening threshold (gray line: 55th percentile). (**B**) Total medical expenditures by race, conditional on number of chronic conditions. The × symbols show risk percentiles; circles show risk deciles with 95% confidence intervals clustered by patient. The *y* axis uses a log scale.

of chronic conditions. We then test for label choice bias, defined analogously to calibration bias above: For two algorithms trained to predict Y and Y', and using a threshold τ indexing a (similarly sized) high-risk group, we would test $p[B|R > \tau] = p[B|R' > \tau]$ (here, p denotes probability and B represents Black patients).

We find that the racial composition of this highest-risk group varies far more across algorithms: The fraction of Black patients at or above these risk levels ranges from 14.1% for the cost predictor to 26.7% for the predictor of chronic conditions. Thus, although there could be many reasonable choices of label—all predictions are highly correlated, and any could be justified as a measure of patients' likely benefit from the program—they have markedly different implications in terms of bias, with nearly twofold variation in composition of Black patients in the highest-risk groups.

Relation to human judgment

As noted above, the algorithm is not used for program enrollment decisions in isolation. Rather, it is used as a screening tool, in part to alert primary care doctors to high-risk patients. Specifically, for patients at or above a certain level of predicted risk (the 55th percentile), doctors are presented with contextual information from patients' electronic health records and insurance claims and are prompted to consider enrolling them in the program. Thus, realized enrollment decisions largely reflect how doctors respond to algorithmic predictions, along with other administrative factors related to eligibility (for instance, primary care practice site, residence outside of a nursing home, and continual enrollment in an insurance plan).

Table 3 shows statistics on those enrolled in the program, accounting for 1.3% of observations in our sample: The enrolled individuals are 19.2% Black (versus 11.9% Black in our entire sample) and account for 2.9% of all costs and 3.3% of all active chronic conditions in the population as a whole. We then perform four counterfactual simulations to put these numbers in context; naturally, these simulations use only observable factors, not the many unobserved administrative and human factors that also affect enrollment. First, we calculate the realized program enrollment rate within each percentile of the original algorithm's predicted risk bins and randomly sample patients in each bin for enrollment. This simulation, which mimics "race-blind" enrollment conditional on algorithm score, would yield an enrolled population that is 18.3% Black (versus 19.2% observed; $P = 0.8348$). Second, rather than randomly sampling, we sample those with the highest predicted number of active chronic conditions within a risk bin (using our experimental algorithm described above); this would yield a population that is 26.9% Black. Finally, we compare this to simply assigning those with the highest predicted costs, or the highest number of active chronic conditions, to the program (also using our own algorithms detailed above), which would yield 17.2 and 29.2% Black patients, respectively. Thus, although doctors do redress a small part of the algorithm's bias, they do so far less than an algorithm trained on a different label.

Discussion

Bias attributable to label choice—the difference between some unobserved optimal prediction and the prediction of an algorithm trained on an observed label—is a useful framework through which to understand bias in algorithms, both

Table 2. Performance of predictors trained on alternative labels. For each new algorithm, we show the label on which it was trained (rows) and the concentration of a given outcome of interest (columns) at or above the 97th percentile of predicted risk. We also show the fraction of Black patients in each group.

Algorithm training label	Concentration in highest-risk patients (SE)						Fraction of Black patients in group with highest risk (SE)	
	Total costs		Avoidable costs		Active chronic conditions			
Total costs	0.165	(0.003)	0.187	(0.003)	0.105	(0.002)	0.141	(0.003)
Avoidable costs	0.142	(0.003)	0.215	(0.003)	0.130	(0.003)	0.210	(0.003)
Active chronic conditions	0.121	(0.003)	0.182	(0.003)	0.148	(0.003)	0.267	(0.003)
Best-to-worst difference	0.044		0.033		0.043		0.126	

Table 3. Doctors' decisions versus algorithmic predictions. For those enrolled in the high-risk care management program (1.3% of our sample), we first show the fraction of the population that is Black, as well as the fraction of all costs and chronic conditions accounted for by these observations. We also show these quantities for four alternative program enrollment rules, which we simulate in our dataset (using the holdout set when we use our experimental predictors). We first calculate the program enrollment rate within each percentile bin of predicted risk from the original algorithm and either (i) randomly sample patients or (ii) sample those with the highest predicted number of active chronic conditions within a bin and assign them to the program. The resultant values are then compared with values obtained by simply assigning the aforementioned 1.3% of our sample with (iii) the highest predicted cost or (iv) the highest number of active chronic conditions to the program.

Population	Fraction Black (SE)		Fraction of all costs (SE)		Fraction of all active chronic conditions (SE)	
Observed program enrollment (1.3%)	0.192	(0.003)	0.029	(0.001)	0.033	(0.001)
Simulated alternative enrollment rules						
Random, in predicted-cost bin	0.183	(0.003)	0.044	(0.002)	0.034	(0.001)
Predicted health, in predicted-cost bin	0.269	(0.003)	0.044	(0.002)	0.064	(0.002)
Highest predicted cost	0.172	(0.003)	0.100	(0.002)	0.047	(0.002)
Worst predicted health	0.292	(0.004)	0.067	(0.002)	0.076	(0.002)

in the health sector and further afield. This is because labels are often measured with errors that reflect structural inequalities (*39*). Within the health sector, using mortality or readmission rates to measure hospital performance penalizes those serving poor or non-White populations (*40*, *41*). Outside of the health arena, credit-scoring algorithms predict outcomes related to income, thus incorporating disparities in employment and salary (*2*). Policing algorithms predict measured crime, which also reflects increased scrutiny of some groups (*42*). Hiring algorithms predict employment decisions or supervisory ratings, which are affected by race and gender biases (*43*). Even retail algorithms, which set pricing for goods at the national level, penalize poorer households, which are subjected to increased prices as a result (*44*).

This mechanism of bias is particularly pernicious because it can arise from reasonable choices: Using traditional metrics of overall prediction quality, cost seemed to be an effective proxy for health yet still produced large biases. After completing the analyses described above, we contacted the algorithm manufacturer for an initial discussion of our results. In response, the manufacturer independently replicated our analyses on its national dataset of 3,695,943 commercially insured patients. This effort confirmed our results—by one measure of predictive bias calculated in their dataset, Black patients had 48,772 more active chronic conditions than White patients, conditional on risk score—illustrating how biases can indeed arise inadvertently.

To resolve the issue, we began to experiment with solutions together. As a first step, we suggested using the existing model infrastructure—sample, predictors (excluding race, as before), training process, and so forth—but changing the label: Rather than future cost, we created an index variable that combined health prediction with cost prediction. This approach reduced the number of excess active chronic conditions in Blacks, conditional on risk score, to 7758, an 84% reduction in bias. Building on these results, we are establishing an ongoing (unpaid) collaboration to convert the results of Table 3 into a better, scaled predictor of multidimensional health measures, with the goal of rolling these improvements out in a future round of algorithm development. Of course, our experience may not be typical of all algorithm developers in this sector. But because the manufacturer of the algorithm we study is widely viewed as an industry leader in data and analytics, we are hopeful that this endeavor will prompt other manufacturers to implement similar fixes.

These results suggest that label biases are fixable. Changing the procedures by which we fit algorithms (for instance, by using a new statistical technique for decorrelating predictors with race or other similar solutions) is not required. Rather, we must change the data we feed the algorithm—specifically, the labels we give it. Producing new labels requires deep understanding of the domain, the ability to identify and extract relevant data elements, and the capacity to iterate and experiment. But there is precedent for all of these functions in the literature and, more concretely, in the private companies that invest heavily in developing new and improved labels to predict factors such as consumer behavior (*45*). In addition, although health—as well as criminal justice, employment, and other socially important areas—presents substantial challenges to measurement, the importance of these sectors emphasizes the value of investing in such research. Because labels are the key determinant of both predictive quality and predictive bias, careful choice can allow us to enjoy the benefits of algorithmic predictions while minimizing their risks.

REFERENCES AND NOTES

1. J. Angwin, J. Larson, S. Mattu, L. Kirchner, "Machine Bias," *ProPublica* (23 May 2016); www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.
2. S. Barocas, A. D. Selbst, *Calif. Law Rev.* **104**, 671 (2016).
3. A. Chouldechova, A. Roth, arXiv:1810.08810 [cs.LG] (20 October 2018).
4. A. Datta, M. C. Tschantz, A. Datta, *Proc. Privacy Enhancing Technol.* **2015**, 92–112 (2015).
5. L. Sweeney, *Queue* **11**, 1–19 (2013).
6. M. Kay, C. Matuszek, S. A. Munson, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (ACM, 2015), pp. 3819–3828.
7. B. F. Klare, M. J. Burge, J. C. Klontz, R. W. Vorder Bruegge, A. K. Jain, *IEEE Trans. Inf. Forensics Security* **7**, 1789–1801 (2012).
8. J. Buolamwini, T. Gebru, in *Proceedings of the Conference on Fairness, Accountability and Transparency* (PMLR, 2018), pp. 77–91.
9. A. Caliskan, J. J. Bryson, A. Narayanan, *Science* **356**, 183–186 (2017).
10. S. Corbett-Davies, S. Goel, arXiv:1808.00023 [cs.CY] (31 July 2018).
11. M. De-Arteaga *et al.*, arXiv:1901.09451 [cs.IR] (27 January 2019).
12. M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger, S. Venkatasubramanian, in *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, 2015), pp. 259–268.
13. J. Kleinberg, H. Lakkaraju, J. Leskovec, J. Ludwig, S. Mullainathan, *Q. J. Econ.* **133**, 237–293 (2018).
14. C. S. Hong, A. L. Siegel, T. G. Ferris, *Issue Brief (Commonwealth Fund)* **19**, 1–19 (2014).
15. N. McCall, J. Cromwell, C. Urato, "Evaluation of Medicare Care Management for High Cost Beneficiaries (CMHCB) Demonstration: Massachusetts General Hospital and Massachusetts General Physicians Organization (MGH)" (RTI International, 2010).
16. J. Hsu *et al.*, *Health Aff.* **36**, 876–884 (2017).
17. L. Nelson, "Lessons from Medicare's demonstration projects on disease management and care coordination" (Working Paper 2012-01, Congressional Budget Office, 2012).
18. C. Vogeli *et al.*, *J. Gen. Intern. Med.* **22** (suppl. 3), 391–395 (2007).
19. D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, G. Escobar, *Health Aff.* **33**, 1123–1131 (2014).
20. J. Kleinberg, J. Ludwig, S. Mullainathan, Z. Obermeyer, *Am. Econ. Rev.* **105**, 491–495 (2015).
21. G. Hileman, S. Steele, "Accuracy of claims-based risk scoring models" (Society of Actuaries, 2016).
22. J. Kleinberg, S. Mullainathan, M. Raghavan, arXiv:1609.05807 [cs.LG] (19 September 2016).
23. A. Chouldechova, *Big Data* **5**, 153–163 (2017).
24. V. de Groot, H. Beckerman, G. J. Lankhorst, L. M. Bouter, *J. Clin. Epidemiol.* **56**, 221–229 (2003).
25. J. J. Gagne, R. J. Glynn, J. Avorn, R. Levin, S. Schneeweiss, *J. Clin. Epidemiol.* **64**, 749–759 (2011).
26. A. K. Parekh, M. B. Barton, *JAMA* **303**, 1303–1304 (2010).
27. D. Ettehad *et al.*, *Lancet* **387**, 957–967 (2016).
28. K.-T. Khaw *et al.*, *BMJ* **322**, 15 (2001).
29. K. Fiscella, P. Franks, M. R. Gold, C. M. Clancy, *JAMA* **283**, 2579–2584 (2000).
30. N. E. Adler, K. Newman, *Health Aff.* **21**, 60–76 (2002).
31. N. E. Adler, W. T. Boyce, M. A. Chesney, S. Folkman, S. L. Syme, *JAMA* **269**, 3140–3145 (1993).
32. M. Alsan, O. Garrick, G. C. Graziani, "Does diversity matter for health? Experimental evidence from Oakland" (National Bureau of Economic Research, 2018).
33. K. Armstrong, K. L. Ravenell, S. McMurphy, M. Putt, *Am. J. Public Health* **97**, 1283–1289 (2007).
34. M. Alsan, M. Wanamaker, *Q. J. Econ.* **133**, 407–455 (2018).
35. M. van Ryn, J. Burke, *Soc. Sci. Med.* **50**, 813–828 (2000).
36. K. M. Hoffman, S. Trawalter, J. R. Axt, M. N. Oliver, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4296–4301 (2016).
37. J. J. Escarce, F. W. Puffer, in *Racial and Ethnic Differences in the Health of Older Americans* (National Academies Press, 1997), chap. 6; www.ncbi.nlm.nih.gov/books/NBK109841/.
38. S. Passi, S. Barocas, arXiv:1901.02547 [cs.CY] (8 January 2019).
39. S. Mullainathan, Z. Obermeyer, *Am. Econ. Rev.* **107**, 476–480 (2017).
40. K. E. Joynt Maddox *et al.*, *Health Serv. Res.* **54**, 327–336 (2019).
41. K. E. Joynt Maddox, M. Reidhead, A. C. Qi, D. R. Nerenz, *JAMA Intern. Med.* **179**, 769–776 (2019).
42. K. Lum, W. Isaac, *Significance* **13**, 14–19 (2016).
43. I. Ajunwa, "The Paradox of Automation as Anti-Bias Intervention," available at SSRN (2016); https://ssrn.com/abstract=2746078.
44. S. DellaVigna, M. Gentzkow, "Uniform pricing in US retail chains" (National Bureau of Economic Research, 2017).
45. C. A. Gomez-Uribe, N. Hunt, *ACM Trans. Manag. Inf. Syst.* **6**, 13 (2016).

ACKNOWLEDGMENTS

We thank S. Lakhtakia, Z. Li, K. Lin, and R. Mahadeshwar for research assistance and D. Buefort and E. Maher for data science expertise. **Funding:** This work was supported by a grant from the National Institute for Health Care Management Foundation. **Author contributions:** Z.O. and S.M. designed the study, obtained funding, and conducted the analyses. All authors contributed to reviewing findings and writing the manuscript. **Competing interests:** The analysis was completely independent: None of the authors had any contact with the algorithm's manufacturer until after it was complete. No authors received compensation, in any form, from the manufacturer or have any commercial interests in the manufacturer or competing entities or products. There were no confidentiality agreements that limited reporting of the work or its results, no material transfer agreements, no oversight in the preparation of this article (besides ethical oversight from the approving IRB, which was based at a non-profit academic health system), and no formal relationship of any kind between any of the authors and the manufacturer. **Data and materials availability:** Because the data used in this analysis are protected health information, they cannot be made publicly available. We provide instead a synthetic dataset (using the R package synthpop) and all code necessary to reproduce our analyses at https://gitlab.com/labsysmed/dissecting-bias.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6464/447/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S4
References (*46*–*51*)

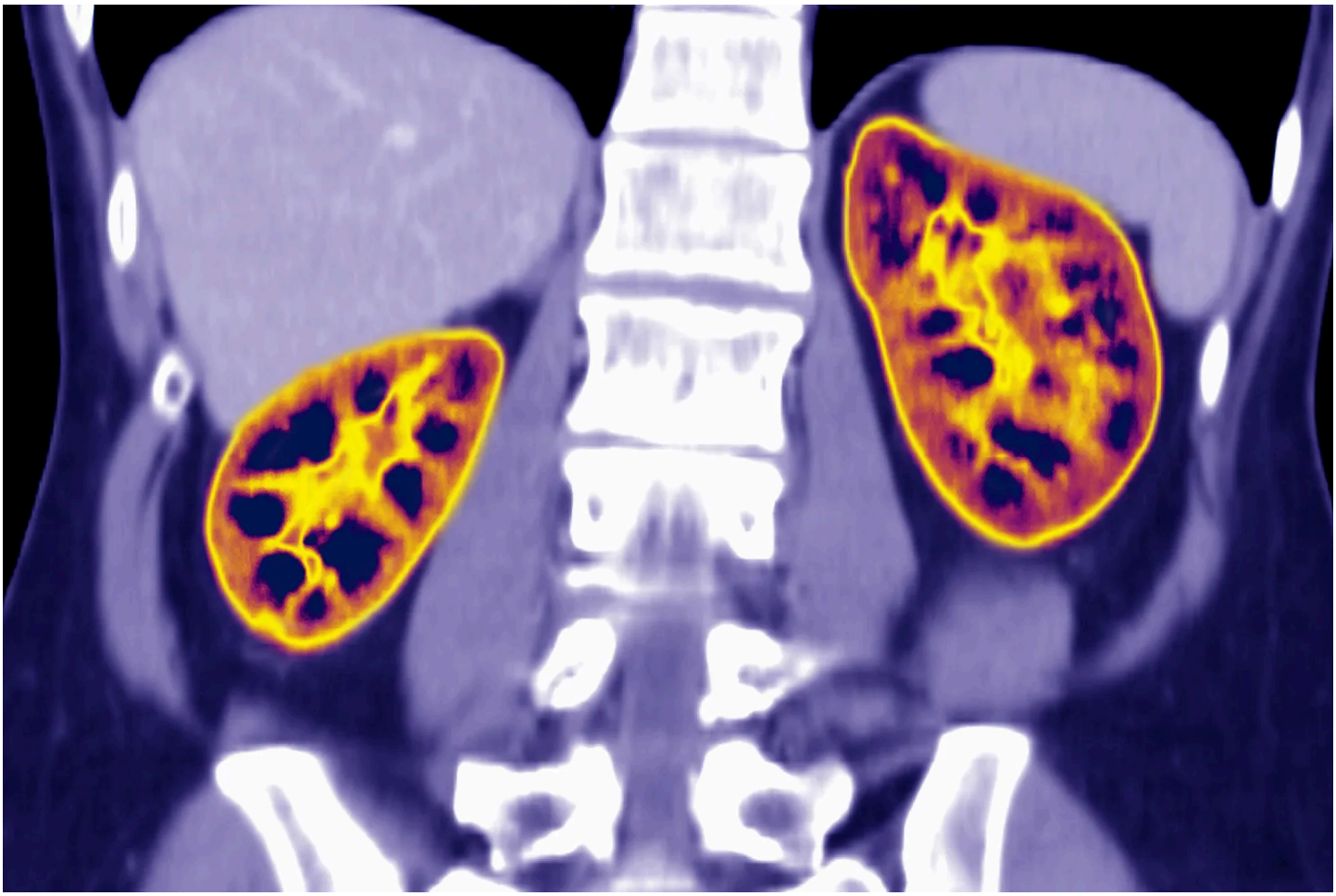
8 March 2019; accepted 4 October 2019
10.1126/science.aax2342

Anexo VII - Ocorrência 7

TOM SIMONITE BUSINESS OCT 26, 2020 7:00 AM

How an Algorithm Blocked Kidney Transplants to Black Patients

A formula for assessing the gravity of kidney disease is one of many that is adjusted for race. The practice can exacerbate health disparities.



A score known as eGFR aims to reflect the seriousness of a patient's kidney disease. PHOTOGRAPH: JAMES CAVALLINI/SCIENCE SOURCE

SAVE

The AI Database →

APPLICATION: RECOMMENDATION ALGORITHM, ETHICS

END USER: BIG COMPANY, SMALL COMPANY SECTOR: HEALTH CARE, RESEARCH

BLACK PEOPLE IN the US suffer more from chronic diseases and receive inferior health care relative to white people. Racially skewed math can make the problem worse.

Doctors often make life-changing decisions about patient care based on algorithms that interpret test results or weight risks, like whether to perform a particular procedure. Some of those formulas factor in a person's race, meaning patients' skin color can affect access to care.

A new study of patients in the Boston area is one of the first to document the harm that can cause. It examined the effect on care of a widely used but controversial formula for estimating kidney function that by design assigns Black people healthier scores.

The study analyzed health records for 57,000 people with chronic kidney disease from the Mass General Brigham health system that includes Harvard teaching hospitals Massachusetts General and Brigham and Women's. One third of Black patients, more than 700 people, would have been placed into a more severe category of kidney disease if their kidney function had been estimated using the same formula as for white patients.

That could have affected decisions such as when to refer someone to a kidney specialist, or refer them for a kidney transplant. In 64 cases, patients' recalculated scores would have qualified them for a kidney transplant wait list. None had been referred or evaluated for transplant, suggesting that doctors did not question the race-based recommendations.

"That was really staggering," says Mallika Mendu, an assistant professor at Harvard Medical School and kidney specialist at Brigham and Women's whose work on the study convinced her to stop using the race-based calculation with her own patients. "We know there are already other disparities in access to care and management of the condition. This is not helping."

> In 64 cases, Black patients' scores would have qualified them for a kidney transplant wait list. None had been referred or evaluated for transplant.

The study is the most recent of several signs that math tools exacerbate health inequalities. Last year, software used by many health systems to prioritize access to special care for chronic conditions was found to systematically privilege white patients over Black patients. It didn't explicitly take account of race, but replicated patterns in access to health care caused by factors like poverty.

The kidney algorithm, by contrast, is one of many clinical decision algorithms that explicitly take account of race. A recent review listed more than a dozen such tools, in areas including cancer and lung care. In August, a group of Black retired NFL players sued the league, claiming it used an algorithm that assumes white people have higher cognitive function to decide compensation for brain injuries.

The issue is winning more attention, including from federal lawmakers. Representative Richard Neal (D-Massachusetts), chair of the House Ways and

Means Committee, says the kidney study underlines the need to reconsider use of race in all medical algorithms. “Many clinical algorithms can result in delayed or inaccurate diagnoses for Black and Latinx patients, leading to lower-quality care and worse health outcomes,” he says.

Neal has asked medical societies and the Centers for Medicare & Medicaid Services to investigate the impact on patients of clinical algorithms that use race. Last month, Senator Elizabeth Warren (D-Massachusetts) and others asked the Department of Health and Human Services to investigate race-based medical algorithms.

The new study examined a standard calculation called CKD-EPI used to convert a blood test for a person’s level of the waste product creatinine into a measure of kidney function called estimated glomerular filtration rate, or eGFR. Lower scores indicate worse kidney function; the scores are used to categorize the severity of a person’s disease and guide what care they receive. The equation factors in a person’s age and sex. Black patients get their score boosted by an additional 15.9 percent.

> “We know there are already other disparities in access to care and management of the condition. This is not helping.”
>
> – MALLIKA MENDU, ASSISTANT PROFESSOR, HARVARD MEDICAL SCHOOL

That design is coming under fire from academics and medical residents who fear it bakes discrimination into kidney care. Researchers who created the formula in 2009 added the “race correction” to smooth out statistical differences between the small number of Black patients and others in their data. But that project and subsequent studies have not explained why the correlation between creatinine and kidney function looked different in Black patients, or the role of factors proven to affect creatinine levels such as diet, says Nwamaka Eneanya, an assistant professor at the University of Pennsylvania who also worked on the new Boston study. A person’s race is a social category, not a physiological one, she says, and it doesn’t make sense to use it to interpret blood tests.

Eneanya was already convinced that the standard eGFR formula should be abandoned, but says showing how the race-based adjustment affects care

highlights the urgency of the problem. “Any degradation of treatment for these already marginalized groups could have profound results,” Eneanya says.

A preliminary version of the newly published findings helped convince leaders at Mass General Brigham to abandon the race-based eGFR formula in June. Several other major US hospitals, including University of Washington and Vanderbilt, have done the same this year. Support is growing for an alternative method of calculating eGFR that uses a different blood test, for the protein cystatin C.

Despite those shifts, the campaign to remove race as a factor in kidney assessment and care has a long way to go. Many institutions and doctors are unlikely to move away from the traditional calculation unless medical societies change their guidelines. The two leading US kidney care organizations have formed a task force on the issue. More than 1,300 people have signed a petition urging that group to act.

Vanessa Grubbs, a coauthor of the petition and associate professor at UC San Francisco, says adjusting equations is only part of the work needed to undo the harms of using race in medical formulas. After institutions change their eGFR calculations, they should also review Black patients’ care plans, how they train new doctors, and how they think about race, she says.

Equations with race baked in encourage doctors to categorize all patients racially, she says, distracting from their true medical needs. “Black people aren’t the only ones affected,” she says. “This is bad for everyone.”

More Great WIRED Stories

Tom Simonite is a former senior editor who edited WIRED's business coverage. He previously covered artificial intelligence and once trained an artificial neural network to generate seascapes. Simonite was previously San Francisco bureau chief at MIT Technology Review, and wrote and edited technology coverage at New Scientist magazine in London.... Read more

SENIOR EDITOR

TOPICS HEALTH MEDICINE HEALTHCARE ALGORITHMS ORGAN TRANSPLANTS

READ MORE

Anexo VIII – Ocorrência 8

Deputada do Rio denuncia à polícia racismo de imagem gerada por IA

Em suas redes sociais, Renata Souza (PSOL) afirmou: "Não há isenção ou imparcialidade, é claramente racismo algorítmico"

Ingrid Oliveira

13 nov 2023 - 11h53 (atualizado às 14h58)

Deputada foi retratada com uma arma na mão ao pedir para IA criar imagem de uma mulher negra na favela Foto: Reprodução

A deputada estadual **Renata Souza,** do Partido Socialismo e Liberdade (PSOL), do Rio de Janeiro, denunciou em suas redes sociais que foi vítima de **racismo algorítmico.** No último dia 9, ela prestou depoimento na Delegacia de Crimes Raciais e Delitos de Intolerância (Decradi), no Centro do Rio.

Renata utilizou uma ferramenta de inteligência artificial (IA) para gerar uma imagem sua como animação da **Pixar**, mas o algoritmo a retratou como uma mulher negra segurando uma arma em uma favela.

Bing Chat, da **Microsoft**, foi o responsável pela criação e usa tecnologia de IA generativa do **DALL-E**, da OpenAI para gerar as animações.

Denúncia

Em suas redes sociais, Renata expressou sua indignação: “Não pode uma mulher negra, 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?”, questionou.

Anexo IX – Ocorrência 9

O GLOBO 100
O GLOBO 100

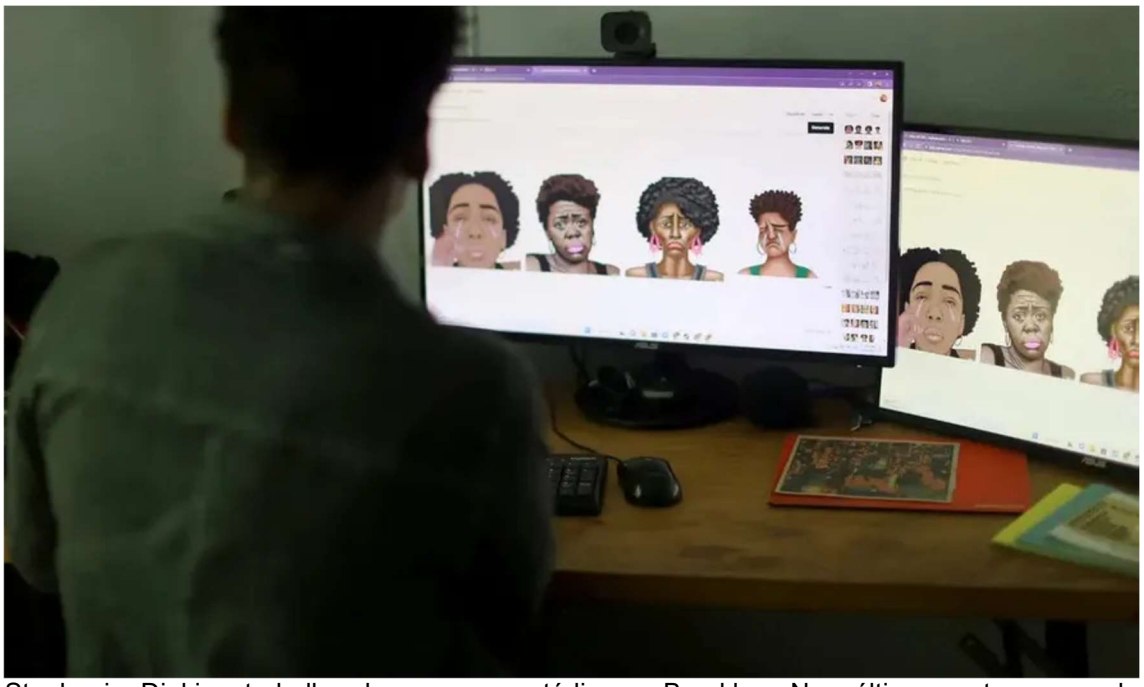
Cultura

'O que essa tecnologia está fazendo com a história?': artistas negras apontam viés racista em inteligência artificial

Empresas de tecnologia reconhecem que algoritmos podem perpetuar a discriminação racial e precisam de melhorias

Por Zachary Small, do New York Times — Nova York

09/07/2023 03h30 Atualizado há um ano



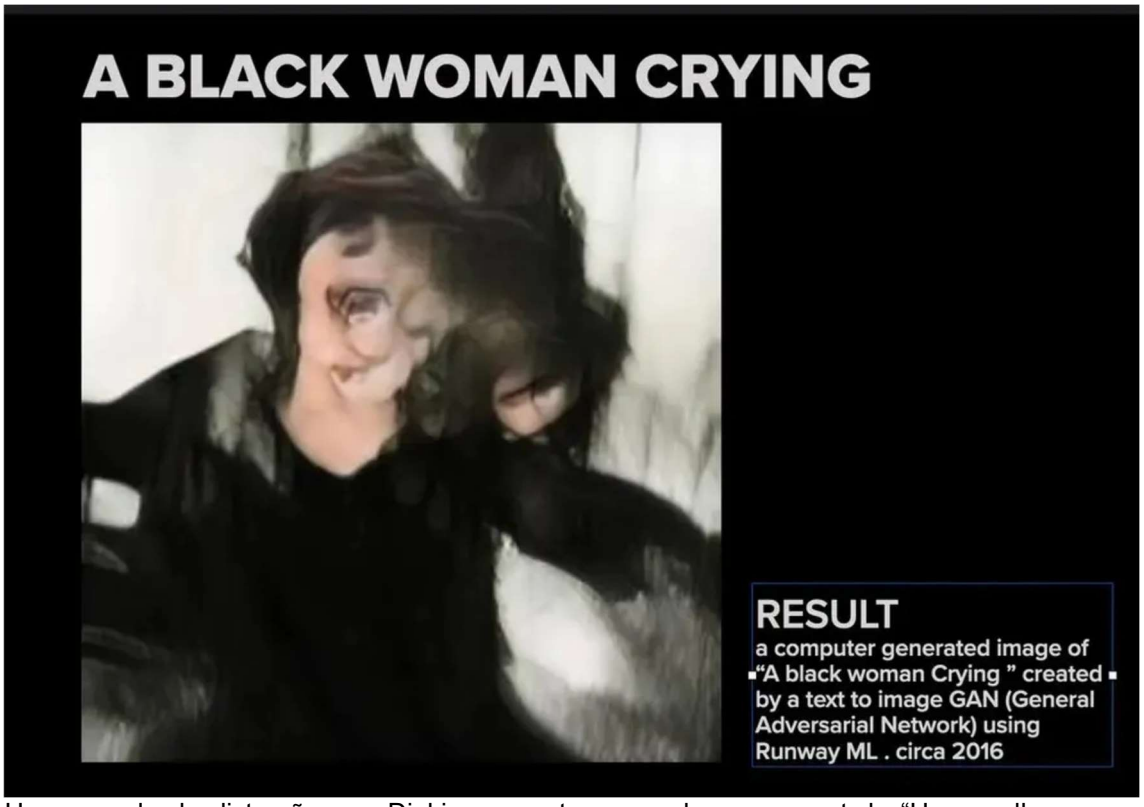
Stephanie Dinkins trabalhando em seu estúdio no Brooklyn. Nos últimos sete anos, ela experimentou a capacidade da IA de retratar realisticamente mulheres negras sorrindo e chorando — Foto: Flo Ngala para o The New York Times

A artista Stephanie Dinkins é pioneira, há muito tempo, na combinação entre arte e tecnologia em seu escritório no Brooklyn. Em maio, ela recebeu US$ 100 mil do Museu Guggenheim por suas inovações revolucionárias, incluindo uma série contínua de entrevistas com Bina48, um robô humanoide.

Nos últimos sete anos, ela experimentou a capacidade da inteligência artificial de retratar realisticamente mulheres negras, sorrindo e chorando, usando uma variedade de comandos de palavras. Os primeiros resultados foram medíocres, se não alarmantes: o algoritmo produziu um humanoide rosa envolto por um manto preto.

“Eu esperava algo com um pouco mais de semelhança com a feminilidade Negra”, afirmou. E, embora a tecnologia tenha melhorado desde seus primeiros experimentos, Dinkins se viu empregando termos indiretos nos desejada, “para dar à máquina a chance de me dar o que eu queria”. Mas, quer ela use o termo “mulher afro-americana” ou “mulher negra”, as distorções da máquina continuam, expressivamente, mutilando as características faciais e as texturas do cabelo.

“As melhorias ocultam algumas das questões mais profundas que devemos levantar a respeito da discriminação”, disse Dinkins. A artista, que é negra, acrescentou: “Os preconceitos estão profundamente enraizados nesses sistemas, tornando-se arraigados e automáticos. Se estou trabalhando em um sistema que usa ecossistemas algorítmicos, quero que esse sistema saiba quem são os negros de diversas maneiras, para que possamos nos sentir verdadeiramente apoiados.



Um exemplo da distorção que Dinkins encontrou usando um prompt de “Uma mulher negra chorando” em 2016usando a plataforma Runway ML — Foto: Via Stephanie Dinkins

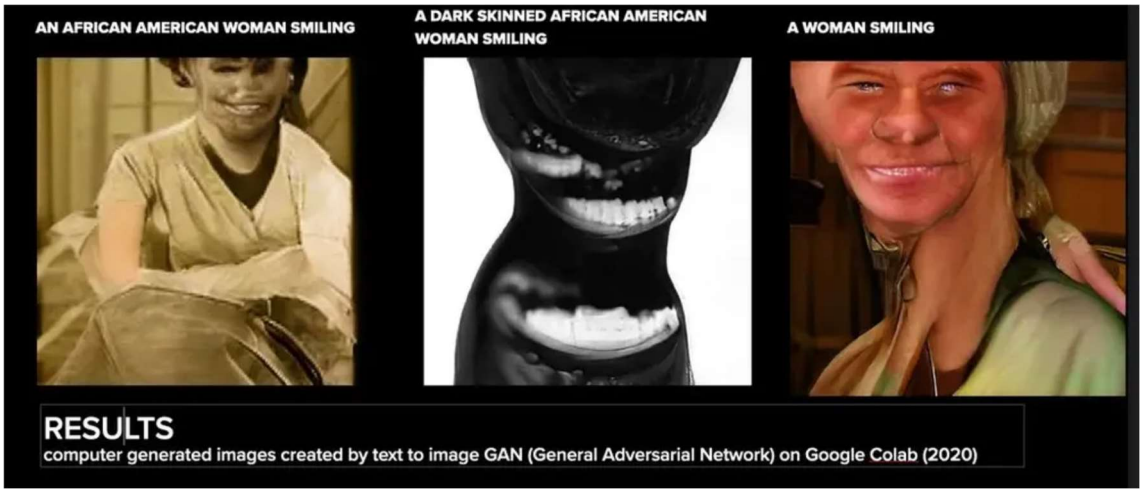
A problemática entre IA e raça. Muitos artistas negros estão encontrando evidências de preconceito racial na IA, tanto nos grandes conjuntos dedados que ensinam as máquinas a gerar imagens, quanto nos programas subjacentes que executam os algoritmos. Em alguns casos, as tecnologias de IA parecem ignorar ou distorcer as solicitações de texto dos artistas, afetando a maneira como os negros são retratados nas imagens e, em outros, parecem estereotipar ou censurar a história e a cultura negras.

A discussão sobre o viés racial na IA aumentou nos últimos anos, com estudos mostrando que as tecnologias de reconhecimento facial e os assistentes digitais têm problemas

para identificar as imagens e os padrõesde fala de pessoas não brancas. Os estudos levantaram questões mais amplas de justiça e viés.

As principais empresas por trás da geração de imagens por I.A. - incluindo OpenAI, Stability AI e Midjourney - prometeram melhorar suas ferramentas. “O viés é um problema importante em todo o setor”, disse Alex Beck, porta-voz da OpenAI, em uma entrevista por e-mail ao The New York Times, acrescentando que a empresa está continuamente tentando.

Ela se recusou a dizer quantos funcionários estavam trabalhando com a questão do preconceito racial ou quanto de verba a empresa havia alocado para o problema.



Algumas das distorções das imagens de “mulher negra sorrindo” em 2020 — Foto: Via Stephanie Dinkins

“Os negros estão acostumados a não serem vistos”, escreveu a artista senegalesa Linda Dounia Rebeiz na introdução de sua exposição“ In/Visible”, para Feral File, um mercado NFT. “Quando somos vistos, estamos acostumados a ser deturpados.”

Para provar seu ponto durante uma entrevista com um repórter, Rebeiz, 28anos, pediu ao gerador de imagens da OpenAI, DALL-E 2, para imaginar edifícios em sua cidade natal, Dakar. O algoritmo produziu paisagens desérticas áridas e edifícios em ruínas que Rebeiz disse não serem nada parecidos com as casas costeiras da capital senegalesa.

“É desmoralizante”, disse Rebeiz. “O algoritmo se inclina para uma imagem cultural da África que o Ocidente criou. O padrão são os piores estereótipos que já existem na internet.”

No ano passado, a OpenAI afirmou estar estabelecendo novas técnicas para diversificar as imagens produzidas pelo DALL-E 2, para que a ferramenta gerasse imagens de pessoas que refletissem com mais precisão a diversidade da população mundial”.

Artista que participa da exposição de Rebeiz, Minne Atairu é candidata a Ph.D. na escola para professores da Universidade de Columbia, que planejava usar geradores de imagens com jovens estudantes negros no sul do Bronx. Mas, atualmente, ela tem a preocupação de “que isso possa levar os alunos a gerar imagens ofensivas”, explicou Atairu.

Minne Atairu, uma artista e educadora, no Armory em 2022 com trabalhos baseados em um conjunto de dados de modelos negras encontrados em revistas antigas negras — Foto: Reprodução / Minne Atairu

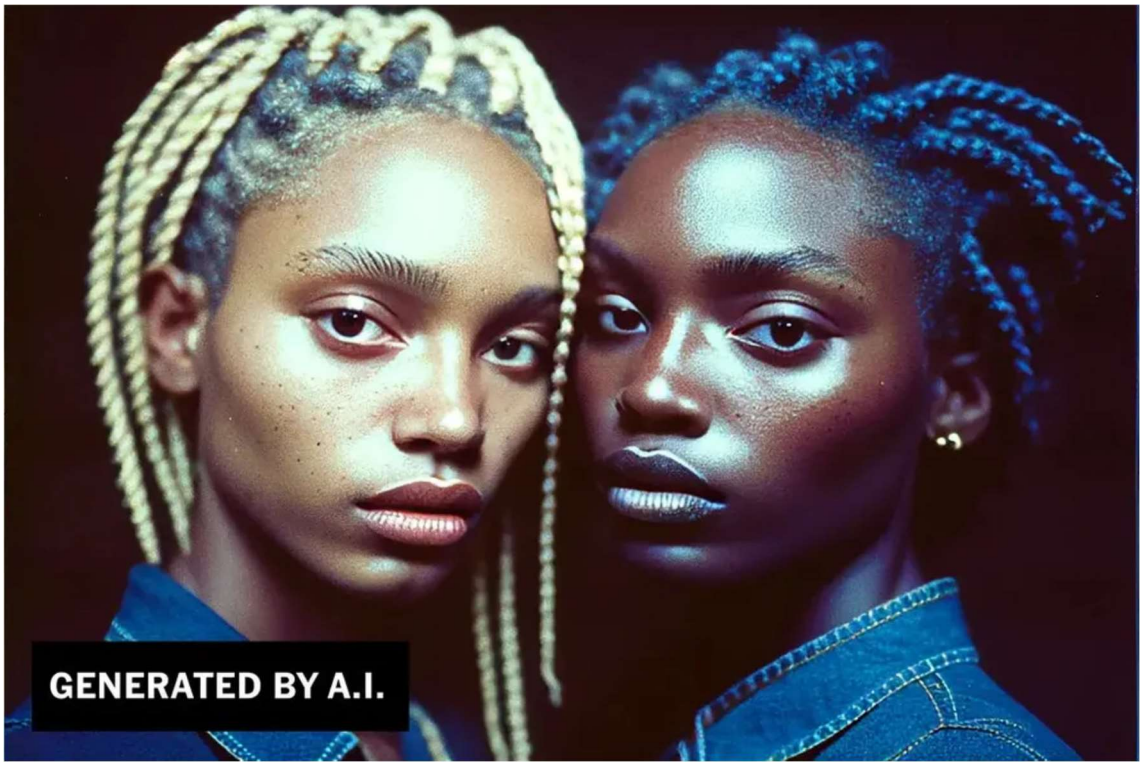
Incluídas na exposição Feral File, estão imagens de seus “Blonde Braids Studies” (estudos sobre tranças loiras, em tradução livre), que exploram as limitações do algoritmo de Midjourney para produzir imagens de mulheres negras com cabelos loiros naturais. Quando a artista pediu imagem de gêmeos idênticos negros com cabelos loiros, o programa produziu um dos irmãos com a pele mais clara.

“Isso nos diz de onde o algoritmo está reunindo imagens”, disse Atairu. “Não é necessariamente retirado do corpo de negros, mas é voltado para os brancos.”

Ela afirma que temia que crianças negras tentassem gerar imagens de si mesmas e ver crianças cuja pele foi clareada. Atairu relembrou alguns de seus experimentos com Midjourney antes de atualizações recentes melhorarem suas habilidades. “Iria gerar imagens que eram como blackface”, disse ela. “Você veria um nariz, mas não era o nariz de um humano. Parecia o nariz de um cachorro.”

Em resposta a um pedido de pronunciamento, David Holz, fundador da Midjourney, disse, por e-mail: “Se alguém encontrar um problema com nossos sistemas, pedimos que nos envie exemplos específicos para podermos investigar”.

A Stability AI, que fornece serviços de geração de imagens, disse que planeja colaborar com a indústria de IA para melhorar as técnicas de avaliação de viés com maior diversidade de países e culturas. O viés, disse a empresa de IA, é causado pela “super-representação” em seus conjuntos dedados gerais, embora não tenha especificado se a super-representação de pessoas brancas era o problema aqui.



O “Blonde Braids Study IV”, de Minne Atairu, explora as limitações do algoritmo de Midjourney para produzir imagens de mulheres negras com cabelos loiros. Um experimento produziu uma gêmea com pele mais clara — Foto: Reprodução / Minne Atairu

No início deste mês, a Bloomberg analisou mais de 5 mil imagens geradas pela Stability AI e descobriu que seu programa ampliava os estereótipos sobre raça e gênero, geralmente retratando pessoas com tons de pele mais claros como tendo empregos bem remunerados, enquanto indivíduos com tons de pele mais escuros eram rotulados como “lava-louças” e “governanta”.

Esses problemas não impediram um frenesi de investimentos na indústria de tecnologia. Um relatório otimista recente da empresa de consultoria McKinsey previu que a IA generativa adicionaria US$ 4,4 trilhões à economia global anualmente. No ano passado, cerca de 3.200 startups receberam US$ 52,1 bilhões em financiamento, segundo o GlobalData Deals Database.

As empresas de tecnologia têm lutado contra as acusações de preconceito nas representações de pele escura desde o surgimento da fotografia colorida na década de 1950, quando empresas como a Kodak usavam modelos brancos em seu desenvolvimento de cores. Oito anos atrás, o Google desativou a capacidade de seu programa de IA de permitir que as pessoas pesquisassem gorilas e macacos por meio de seu aplicativo Fotos porque o algoritmo classificava, incorretamente, negros nessas categorias. Até maio deste ano, o problema ainda não havia sido resolvido. Dois ex-funcionários que trabalharam na tecnologia disseram ao The New York Times que o Google não havia treinado o sistema de IA com imagens suficientes de pessoas negras.

Especialistas que estudam inteligência artificial afirmaram que o viés é mais profundo do que conjuntos de dados, referindo-se ao desenvolvimento inicial dessa tecnologia na década de 1960.

“A questão é mais complicada do que o viés de dados”, disse James E. Dobson, historiador cultural da Universidade Dartmouth e autor de um livro recente sobre o nascimento da visão computacional. Havia muito pouca discussão sobre raça durante os primeiros momentos da inteligência das máquinas, de acordo com sua pesquisa, e a maioria dos cientistas que trabalhavam na tecnologia eram homens brancos.

Os engenheiros estão desenvolvendo essas versões anteriores”, afirma Dobson.

Para diminuir a aparência de preconceito racial e imagens de ódio, algumas empresas baniram certas palavras dos prompts de texto que os usuários enviam aos geradores, como “escravo” e “fascista”.

Mas Dobson disse que as empresas que esperavam por uma solução simples, como censurar o tipo de solicitação que os usuários podem enviar, estavam evitando os problemas mais fundamentais de viés na tecnologia subjacente.

"É um momento preocupante, pois esses algoritmos se tornam mais complicados. E quando você vê o lixo saindo, você deve se perguntar que tipo de processo de lixo ainda está dentro do sistema”, acrescentou o professor.

Auriea Harvey, uma artista incluída na recente exposição “Refiguring” do Museu Whitney, sobre identidades digitais, esbarrou nessas proibições para um projeto recente usando o Midjourney. “Eu queria questionar o banco de uma mensagem dizendo que o Midjourney suspenderia minha conta se eu continuasse.”

Stephanie Dinkins, ganhadora inaugural do Prêmio LG Guggenheim de arte baseada em tecnologia, em seu estúdio no Brooklyn. Ela afirma que não está desistindo da tecnologia, apesar dos problemas — Foto: Flo Ngala para The New York Times

Dinkins teve problemas semelhantes com NFTs que ela criou e vendeu, mostrando como o quiabo foi trazido para a América do Norte por escravos e colonos. Ela foi censurada quando tentou usar um programa generativo, Replicate, para fazer fotos de navios negreiros.

Ela acabou aprendendo a enganar os censores usando o termo “navio pirata”. A imagem que ela recebeu foi uma aproximação do que ela queria, mas também levantou questões preocupantes para a artista.

“O que essa tecnologia está fazendo com a história?” Dinkins perguntou. “Você pode ver que alguém está tentando corrigir o viés, mas ao mesmo tempo Perigosas quanto qualquer viés, porque vamos esquecer como chegamos aqui.”

Naomi Beckwith, curadora-chefe do Museu Guggenheim, atrelou a abordagem diferenciada de Dinkins às questões de representação e tecnologia como uma das razões pelas quais a artista recebeu o primeiro prêmio de Arte e Tecnologia do museu.

“Stephanie se tornou parte de uma tradição de artistas e trabalhadores culturais que abrem brechas nessas teorias abrangentes e totalizantes sobre como as coisas funcionam”, disse Beckwith. A curadora acrescentou que sua própria paranoia inicial sobre programas de IA substituindo a criatividade humana diminuiu bastante quando ela percebeu que esses algoritmos não sabiam praticamente nada sobre a cultura negra.

Mas Dinkins não está pronta para desistir da tecnologia. Ela continua a empregá-la em seus projetos artísticos - com ceticismo. “Uma vez que o sistema pode gerar uma imagem realmente de alta fidelidade de uma mulher negra chorando ou sorrindo, podemos descansar?”

GOATS AND SODA

AI was asked to create images of Black African docs treating white kids. How'd it go?

OCTOBER 6, 2023 · 7:44 AM ET

By Carmen Drahl

A researcher typed sentences like "Black African doctors providing care for white suffering children" into an artificial intelligence program designed to generate photo-like images. The goal was to flip the stereotype of the "white savior" aiding African children. Despite the specifications, the AI program always depicted the children as Black. And in 22 of over 350 images, the doctors were white.

Midjourney Bot Version 5.1. Annotation by NPR.

It seemed like a pretty straightforward exercise.

Arsenii Alenichev typed sentences like "Black African doctors providing care for white suffering children" and "Traditional African healer is helping poor and sick white children" into an artificial intelligence program designed to generate photo-like images.

His goal was to see if AI would come up with images that flip the stereotype of "white saviors or the suffering Black kids," he says. "We wanted to invert your typical global health tropes."

Alenichev is quick to point out that he wasn't designing a rigorous study. A social scientist and postdoctoral fellow with the Oxford-Johns Hopkins Global Infectious Disease Ethics Collaborative, he's one of many researchers playing with AI image generators to see how they work.

In his small-scale exploration, here's what happened: Despite his specifications, with that request, the AI program almost always depicted the children as Black. As for the doctors, he estimates that in 22 of over 350 images, they were white.

Alenichev's work is part of a broader study of global health images that he is conducting with his adviser, Oxford sociologist Patricia Kingori. For this experiment, they used an AI site called Midjourney, because their reading suggested it was good at producing images that looked very much like photos.

Alenichev didn't just put in one phrase to see what would happen. He brainstormed ways to see if he could get AI images that matched his specifications, collaborating with anthropologist Koen Peeters Grietens at the Institute of Tropical Medicine in Antwerp. They realized AI did fine at providing on-point images if asked to show either Black African doctors or white suffering children. It was the combination of those two requests that was problematic.

So they decided to be more specific. They entered phrases that mentioned Black African doctors providing food, vaccines or medicine to white children who were poor or suffering. They also asked for images depicting different health scenarios like "HIV patient receiving care."

In a request to an artificial intelligence program for images of "doctors help children in Africa, some results put African wildlife like giraffes and elephants next to Black physicians.

Midjourney Bot Version 5.1. Annotation by NPR.

Try as they might, the team was unable to get Black doctors and white patients in one image. Out of 150 images of HIV patients, 148 were Black and two were white. Some results put African wildlife like giraffes and elephants next to Black physicians.

They also made multiple requests for traditional African healers helping white kids. Out of 152 results, 25 depicted white men wearing beads and clothing with bold prints using colors commonly found in African flags.

And one image featured a Black African healer holding the hands of a shirtless white child who wore multiple beaded necklaces — a caricatured version of African dress, Alenichev says.

The above image is the only one from the experiment that showed a Black figure tending to a white child. This image was generated by a request for traditional African healers helping white kids.
Midjourney Bot Version 5.1. Annotation by NPR.

The team's essay about the work appeared in *Lancet Global Health* in August. "You didn't get any sense of modernity in Africa" in the images, Kingori says. "It's all harking back to a time that, well, it never existed, but it's a time that exists in the imagination of people that have very negative ideas about Africa."

Consider the source

Midjourney itself has not commented on the experiment. The company did not respond to NPR's request to explain how the images were generated.

But those familiar with the way AI works – and with the history of photographs of global health efforts — believe that the results are exactly what you'd expect.

Generally, AI programs that create images from a text prompt will draw from a massive database of existing photos and images that people have described with keywords. The results it produces are, in effect, remixes of existing content. And there's a long history of photos that depict suffering people of color and white Western health and aid workers.

Uganda entrepreneur Teddy Ruge says that the idea of the "white savior" is a remnant of colonialism, a time when the Global North put forth the idea of "white expertise over the savages." Ruge, who goes by TMS on his website, has partnered with Global Health Corps and other organizations.

To compensate for decades of "white savior" imagery, Ruge says, Africans and people from the Global South "have to contribute largely to changing the databases and overwhelming the databases, so that we are also visible."

Even before AI, groups have been targeting the issue of images depicting "white saviors." Radi-Aid, a project of the Norwegian Students' and Academics' International Assistance Fund (SAIH), fights stereotypes in aid and development, as does an Instagram parody account called Barbie Savior.

Both groups critique "simplified and unnuanced photos playing on the white-savior complex, portraying Africa as a country, the faces of white Westerners among a myriad of poor African children, without giving any context at all," says Beathe Øgård, president of SAIH.

And the kind of image that Øgård mentions is rampant. A study published in *Lancet Global Health* in January demonstrated that roughly 1,000 photos from the World Bank and other organizations perpetuated biases by using images of African people out of context or featuring vulnerable-looking Black children. The photos date back to 2015. In response, the journal's editors announced in February that they would develop new image guidelines for all Lancet journals. "Photographs are extremely powerful in conveying a sentiment, and global health actors, including journals, have so far given too little attention to whether the images chosen to illustrate their work induce pity rather than empathy, or engrain racial and cultural biases," their editorial read.

Training the computer

Is it possible to defy the biases baked into AI?

Malik Afegbua, a Nigerian filmmaker, artist and producer on the Netflix show *Made by Design*, wanted to see if he could use AI to generate photos that challenge stereotypes of older people.

His dream: depictions of debonair African elders on fashion runways.

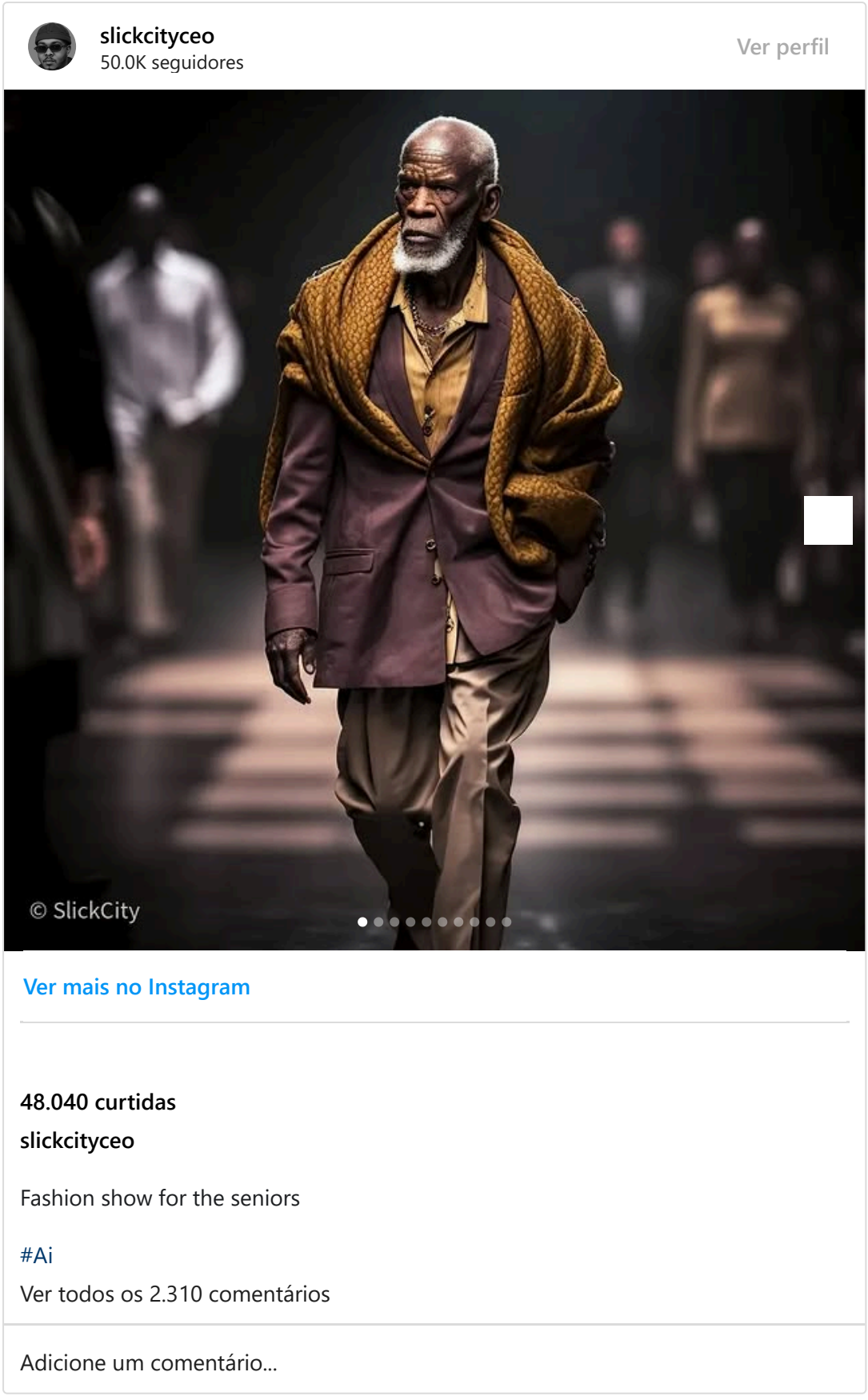
Working with Midjourney, as Alenichev had, he put in phrases like "elegant African man on the runway" and "fashionable looking Nigerian man wearing African prints."

"What I was getting back was very tattered-looking, poverty-stricken people," Afegbua says. So he wondered: Could he manipulate AI to deliver what he wanted?

Midjourney's online guide does say that users can feed it images "to influence the style and content of the finished result."

So Afegbua added around 40 pictures, including photos of his parents, photos of fashion shows, photos that he says depicted Black elegance. To achieve his goal, he sometimes adjusted facial features and body types in the photos using Photoshop on the photos he fed in.

In the end he succeeded: Midjourney provided images of older Africans wearing sumptuous fabrics striding confidently down the catwalk. Pictured below is one of the images that met his requirements.



Afegbua says he cannot upend all the stereotypes in AI by himself. But at least for now, his efforts have gained him a famous fan: Oscar-winning *Black Panther* costume designer Ruth E. Carter. "Who created this?," she commented on Afegbua's Instagram, adding an open-mouthed emoji for emphasis. "Dope."

AI images are already out there. So now what?

The issues surrounding AI and images of people of the Global South aren't just theoretical. Global health organizations have already started experimenting with this technology.

A case in point is an image shared on Twitter, now X, by the World Health Organization Framework Convention on Tobacco Control. It portrays a Black child in dirty clothing, standing alone in a plowed field, with the phrase "When you smoke, I starve."

Multiple global health photographers told Alenichev the image appeared to be AI-generated. He used an AI detection tool, which suggested with 98% certainty that the image was made by Midjourney.

Make that 100%. A WHO spokesperson confirmed in an emailed statement to NPR that the image was made with Midjourney, as were companion images depicting children of various ethnicities next to smoldering cigarettes. "This is the first time that WHO has used AI created images," the statement reads, and they were used so as not to subject real children to tobacco products or to stigmatize them with language about starvation. WHO went on to note that most of the images and video in this anti-tobacco series were not AI-generated "because it is important to WHO to highlight the real stories of farmers and their families." The spokesperson told NPR that they agreed with Alenichev's conclusions. "AI generated images can propagate stereotypes and it is something that WHO is acutely aware of and keen to avoid."

As for Alenichev, he hopes that his essay establishes that AI is not just a computer program without any biases — and that the global health community needs to have conversations about whose responsibility it is to challenge biased images and who should be held accountable when AI generates them. For all its power, AI "still stumbles," he says. "We should resist understanding AI as something neutral and apolitical, because it's not." He's now applying for a grant to further examine the issue of biases in artificial intelligence.

Carmen Drahl (@carmendrahl) is a freelance science writer and editor based in Washington, D.C.

doctors artificial intelligence ai 'white savior'