

## Instâncias textuais do racismo algorítmico

**JÚLIO ARAÚJO**  
araujo@ufc.br

Doutor em Linguística pela Universidade Federal do Ceará, Brasil (2006)  
Professor Titular da Universidade Federal do Ceará, Brasil  
Bolsista de Produtividade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brasil) (nº do processo: 303677/2025-2)

**PALAVRAS-CHAVE:**  
Racismo Algorítmico;  
Necroalgoritmização;  
Inteligência Artificial;  
Análise Discursiva.

**ARTIGO RECEBIDO A:**  
15/08/2025  
**ARTIGO APROVADO A:**  
17/11/2025

**RESUMO:** Neste artigo, examino algoritmos de inteligência artificial (IA) como textos estruturantes que processam informações, moldam narrativas e reproduzem estereótipos raciais, sustentando dinâmicas de exclusão. Apresento um modelo teórico com três instâncias textuais interligadas: Texto-Prompt, que expressa a intenção do usuário; Texto-Resposta, que materializa o resultado produzido; e Texto-Algoritmo, camada invisível que define as regras internas do sistema. Analiso um caso em que uma imagem gerada por IA retratou uma mulher negra armada em uma favela, embora o pedido original indicasse uma figura positiva. Para explorar esse viés, desenvolvi uma simulação em Python que evidencia a associação automática entre termos racializados e violência, revelando a reprodução sistemática de estereótipos. Também examino comentários em redes sociais que naturalizam essas distorções, transferindo a responsabilidade ao usuário e ocultando o papel central do algoritmo na construção das representações. Sustento que tais sistemas funcionam como agentes discursivos que reforçam desigualdades e hierarquias históricas. Defendo a reescrita algorítmica como gesto ético e político, apoiada em estratégias como diversificação de bases de dados, filtros de viés, transparência e regulamentação. Concluo reafirmando a tese que orienta este trabalho: o algoritmo é um texto e, como tal, pode e deve ser reescrito para romper lógicas de exclusão.



## Textual Instances of Algorithmic Racism

JÚLIO ARAÚJO  
araujo@ufc.br

PhD in Linguistics from the Universidade Federal do Ceará, Brazil (2006)  
Full Professor at the Universidade Federal do Ceará, Brazil  
Productivity Research Fellow of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) (Process No. 303677/2025-2)

**KEYWORDS:**

Algorithmic Racism;  
Necroalgorithmization;  
Artificial Intelligence;  
Discourse Analysis.

**ARTICLE RECEIVED ON:**  
08/15/2025

**ARTICLE ACCEPTED ON:**  
11/17/2025

**ABSTRACT:** In this article, I examine artificial intelligence (AI) algorithms as structuring texts that process information, shape narratives, and reproduce racial stereotypes, thereby sustaining dynamics of exclusion. I present a theoretical model composed of three interconnected textual instances: the Prompt-Text, which expresses the user's intent; the Response-Text, which materializes the generated output; and the Algorithm-Text, the invisible layer that defines the system's internal rules. I analyze a case in which an AI-generated image depicted a Black woman holding a gun in a favela, despite the original prompt specifying a positive figure. To explore this bias, I developed a Python simulation that demonstrates how the algorithm automatically associates racialized terms with violence, revealing the systematic reproduction of stereotypes. I also examine social media comments that naturalize such distortions, shifting responsibility to the user while obscuring the algorithm's central role in constructing representations. I argue that these systems operate as discursive agents that reinforce historical inequalities and hierarchies. I defend algorithmic rewriting as an ethical and political act, supported by strategies such as diversifying datasets, applying bias filters, ensuring greater transparency, and promoting regulation. I conclude by reaffirming the thesis that guides this work: the algorithm is a text, and, as such, it can – and must – be rewritten to dismantle exclusionary logics.



## INTRODUÇÃO

A crescente presença da inteligência artificial (IA) em processos de produção de conteúdo tem evidenciado a influência dos algoritmos na construção de narrativas e significados sociais. Longe de serem entidades neutras e objetivas, os algoritmos carregam, em sua estrutura, decisões humanas, baseadas em dados historicamente condicionados por relações de poder e desigualdade. Assim, em vez de operarem como meras ferramentas técnicas, os algoritmos configuram-se como textos (Araújo, 2025a) e, em função disso, funcionam como construções discursivas que organizam e reproduzem sentidos, afetando a maneira como determinados grupos sociais são representados.

No Brasil, esse fenômeno tornou-se ainda mais evidente com o recente caso da deputada estadual Renata Souza (PSOL-RJ), que utilizou uma ferramenta de IA para gerar uma imagem sua no estilo Pixar. Embora o prompt descrevesse apenas sua identidade como mulher negra e sua origem na favela da Maré, a IA produziu uma imagem que a retratava segurando uma arma, ativando uma narrativa visual associada à criminalidade. Esse episódio ilustra como sistemas algorítmicos não interpretam comandos de maneira literal e mobilizam padrões e associações estatísticas presentes nos dados de treinamento, reproduzindo estereótipos historicamente arraigados. A forma como diferentes identidades são visualmente representadas nesses sistemas reflete, portanto, hierarquias sociais e construções discriminatórias já existentes.

Para leitoras e leitores não brasileiros, é fundamental contextualizar quem é a figura pública envolvida no episódio que desencadeou este estudo. Renata Souza é deputada estadual no Rio de Janeiro e uma importante intelectual negra brasileira. Graduada em Comunicação Social pela Universidade Federal do Rio de Janeiro (UFRJ), possui mestrado e doutorado em Co-

municação e Cultura pela mesma instituição, além de pós-doutorado em Comunicação pela Universidade Federal Fluminense (UFF). Sua trajetória acadêmica e política articula a defesa dos direitos humanos, a denúncia do genocídio da população negra e a crítica sistemática da violência de Estado nas favelas cariocas. Diante desse perfil, o fato de um sistema de IA generativa produzir uma imagem em que a parlamentar aparece segurando uma arma evidencia, de maneira ainda mais contundente, como essas tecnologias podem automatizar e atualizar estereótipos raciais profundamente enraizados na história social brasileira.

Este artigo demonstra que, por serem textos, os algoritmos produzem narrativas visuais capazes de reforçar estereótipos raciais. A partir da análise do caso de Renata Souza, sustenta-se que, quando um prompt contém termos racializados como “mulher negra”, “afro” ou “favela”, o sistema tende a associá-los automaticamente a elementos que reforçam narrativas de violência e criminalidade. Em contraste, prompts sem essas referências resultam em imagens mais positivas, com representações distantes da criminalidade. Esse padrão revela que os algoritmos geram imagens e constroem discursos visualmente orientados, capazes de reproduzir ou ressignificar narrativas sociais preexistentes.

A tese defendida neste artigo é a de que esses processos de geração de imagens por IA demonstram que o algoritmo não opera de forma neutra, mas como um texto passível de reescrita. Tal como ocorre com qualquer produção discursiva, os algoritmos podem e devem ser reconfigurados para evitar a reprodução de vieses discriminatórios. A compreensão dessa dinâmica é fundamental para o desenvolvimento de tecnologias mais justas, capazes de promover representações equânimes e de contrariar a lógica de exclusão automatizada que caracteriza a necroalgoritmização (Araújo, 2025b).

Por meio desta investigação, evidencia-se os mecanismos de reprodução do racismo algorítmico e a urgência de uma intervenção crítica nesses sistemas. Se o algoritmo é um texto, permanece a pergunta: a quem cabe reescrevê-lo?

## 1. A NATUREZA DISCURSIVA DOS ALGORITMOS

### 1.1. O QUE É UM ALGORITMO?

Robin K. Hill (2016), em “What an Algorithm Is”, argumenta que algoritmos não são meras instruções mecânicas, mas construções formais que carregam significados e pressuposições. Para ela, a definição envolve aspectos sintáticos e semânticos, pois, além de organizar comandos, incorpora sentidos manifestos em suas funções e impactos. Assim, um algoritmo estrutura operações e reflete valores, normas e interpretações de quem o escreve.

Na perspectiva da Filosofia da Computação, Hill propõe uma definição que ultrapassa o formalismo e enfatiza estrutura, controle e propósito: “um algoritmo é uma estrutura de controle composta, finita, abstrata e efetiva, dada de forma imperativa, que realiza um propósito específico sob condições previamente estabelecidas” (Hill, 2016, p. 47)<sup>1</sup>. Essa concepção amplia a visão tradicional ao atribuir-lhe status ontológico, reconhecendo-o como entidade estruturada que impõe regras e opera dentro de condições específicas. A autora identifica seis características essenciais: finito, pois deve caber em espaço e tempo delimitados; abstrato, por não se restringir a manifestação física e ser aplicável em diferentes contextos; efetivo, por ser executável sem depender de julgamento humano ou criatividade; estruturado, por organizar operações em sequência lógica; imperativo, por prescrever procedimentos de forma explícita e não ambígua; e intencional, por visar a um propósito específico em condições determinadas.

1. Tradução livre de: “An algorithm is a finite, abstract, effective, compound control structure, imperatively given, accomplishing a given purpose under given provisions” (Hill, 2016, p. 47).

2. Tradução livre de: “there is way more to an algorithm than its procedure” (Hill, 2016, p. 46).

3. Tradução livre de: “when we teach algorithms, we do not merely hand our students Turing machines... we teach an algorithm via vernacular, hand-waving and demonstration, until I see the light dawn” (Hill, 2016, p. 54).

4. Tradução livre de: “algorithms are all about semantics, with ‘what it means’ taken to be ‘what it does’ for us” (Hill, 2016, p. 46).

A partir dessa definição, defende-se que “o algoritmo é um texto” (Araújo, 2025a), pois, como sistema de significado, processa, organiza e interpreta dados. Hill (2016, p. 46) afirma que “há muito mais em um algoritmo do que apenas seu procedimento”<sup>2</sup>, ressaltando que interpretação e contexto de uso são fundamentais à sua existência. Tal como ocorre com discursos, estruturam narrativas e moldam percepções da realidade ao definir a organização e a apresentação de dados, gerando efeitos de sentido alinhados às intenções de seus programadores.

A natureza textual se evidencia quando Hill observa que a compreensão de um algoritmo não se reduz à codificação formal, pois “quando ensinamos algoritmos, não apenas entregamos aos nossos alunos máquinas de Turing... ensinamos um algoritmo por meio da linguagem, gestos e demonstração”<sup>3</sup> (Hill, 2016, p. 54). Isso confirma que não existem em um vácuo técnico, mas são explicados, interpretados e implementados em contextos que influenciam sua aplicação e constroem sentidos.

Ao reconhecer que operam dentro de uma estrutura discursiva, torna-se evidente a impossibilidade de considerar tais sistemas como neutros. Hill (2016) observa que a definição clássica frequentemente ignora a dimensão semântica, tratando os algoritmos apenas como fórmulas matemáticas, quando “os algoritmos são essencialmente sobre semântica, com ‘o que significa’ sendo entendido como ‘o que faz’ para nós”<sup>4</sup> (Hill, 2016, p. 46). Assim, os algoritmos carregam valores, pressupostos e intenções incorporados por seus criadores, refletindo escolhas que podem amplificar determinados padrões e invisibilizar outros, dinâmica especialmente crítica em aplicações como avaliação de crédito, seleção de candidatos e policiamento preditivo.

O modo como algoritmos processam dados afeta também a produção de representações visuais. O recente trend de imagens geradas por IA no estilo Pixar, por exemplo, estiliza traços

físicos e contextos sociais a partir de bases enviesadas. Esse processo pode reforçar estereótipos e distorcer narrativas, como no caso denunciado pela deputada Renata Souza, cuja imagem foi associada a um cenário de violência, expondo as implicações do racismo algorítmico na geração de imagens (Araújo, 2024).

Essa discussão se articula com a análise de Hill sobre como a definição e o ensino de algoritmos influenciam sua implementação e seu impacto social. A autora questiona quem decide quais algoritmos merecem estudo e aplicação e quais problemas são considerados dignos de solução algorítmica, apontando que tais escolhas orientam os critérios e as prioridades do desenvolvimento tecnológico.

A abordagem de Hill revela que algoritmos não são apenas estruturas matemáticas, mas discursos que moldam o mundo digital e social. Reconhecer sua dimensão semântica e interpretativa reforça a tese de que são textos que carregam intenções, vieses e efeitos concretos. Assim, analisá-los exige ir além da eficiência técnica, incorporando sua natureza discursiva, política e ética. Essa perspectiva é fundamental para compreender que, embora escritos para um fim específico, algoritmos são interpretados de formas que podem perpetuar estruturas preexistentes e veicular ideologias, influenciando a forma como processam dados e produzem resultados.

## **1.2. O MITO DA NEUTRALIDADE ALGORÍTMICA**

A crença na neutralidade algorítmica é um dos grandes equívocos da nossa era. O discurso da imparcialidade matemática oculta que tais sistemas são produtos de decisões humanas, carregando valores, vieses e interesses. Como alerta Cathy O’Neil, “os algoritmos são ‘opiniões incorporadas em matemática’” (O’Neil, 2020, p. 15), evidenciando que modelos computacionais não são abstrações neutras, mas escolhas que frequentemente reproduzem desigualdades sistêmicas.

O’Neil mostra como algoritmos aplicados em recrutamento, crédito e segurança pública operam em prejuízo de populações marginalizadas. Seu argumento central destaca três características: opacidade, escala e dano. Destaco a opacidade para mostrar que ela decorre da impossibilidade de compreender e contestar decisões algorítmicas, muitas vezes protegidas como “segredos comerciais vitais” (O’Neil, 2020, p. 208).

A nocividade desses sistemas se amplia no campo da busca e recomendação, como analisa Safiya Noble (2021). Segundo a autora, “o racismo não é um fenômeno novo e, cada vez mais, as formas como ele se manifesta vão se alterando e se aprimorando. Hoje, ele tem se perpetuado também no campo das tecnologias digitais” (Noble, 2021, p. 14). Noble examina o Google e revela como estruturas de ranqueamento favorecem representações estereotipadas e discriminatórias da mulher negra, mantendo hierarquias de poder sob a ilusão de neutralidade.

O efeito combinado dessas dinâmicas configura a exclusão algorítmica, fenômeno no qual grupos historicamente subjugados são sistematicamente prejudicados por decisões automatizadas. Os algoritmos “não ouvem. Tampouco vergam. São surdos não somente a charme, ameaça e adulação, mas também à lógica – mesmo quando há boas razões para se questionar os dados que alimentam seus resultados” (O’Neil, 2020, p. 18). Essa insensibilidade alimenta um círculo vicioso em que erros deixam de ser pontuais e se tornam estruturais. Reconhecer tais impactos exige desmontar o mito da neutralidade e admitir que algoritmos operam dentro de estruturas sociais marcadas por desigualdades. Para Noble, “é hora de pensar sobre inteligência artificial como uma questão de direitos humanos” (Noble, 2021, p. 59). A desmistificação da imparcialidade matemática é condição para reivindicar transparência, regulamentação e design comprometido com justiça social.

Essa discussão sobre exclusão algorítmica e sobre a necessidade de desnaturalizar a pretensa neutralidade técnica encaminha para um ponto central das pesquisas recentes sobre IA generativa: sua natureza eminentemente linguística. Como observam Duque-Pereira e Moura (2023), a IA generativa deve ser compreendida como um fenômeno linguístico que reorganiza práticas comunicativas, uma vez que opera por meio de simulações estatísticas e não por processos interpretativos. Os autores ressaltam a opacidade dos modelos e os riscos associados a corpora de treinamento enviesados, o que reforça a necessidade de abordagens críticas que considerem os efeitos sociopolíticos dessas arquiteturas. Essa reflexão converge com a perspectiva aqui adotada de que o algoritmo é também um texto performativo.

Essa crítica converge com o conceito de governamentalidade algorítmica de García-Canclini (2021), para quem os algoritmos “desorganizam massas” e, sob aparência de neutralidade, reforçam desigualdades ao excluir grupos da visibilidade digital. O autor adverte que “a invisibilidade gerada pelos algoritmos é um perigo real, pois, quando os dados falham em representar a diversidade das experiências humanas, reforçam a exclusão dos grupos marginalizados” (García Canclini, 2019, p. 149).

O debate também dialoga com o colonialismo digital de Faustino e Lippold (2023), forma concreta de dominação que perpetua desigualdades econômicas, políticas e raciais por meio da tecnologia. Treinados com dados enviesados, os algoritmos atualizam a lógica colonial sob o verniz da inovação. Essa lógica se radicaliza no conceito de necroalgoritmização, que examina como a tecnologia hierarquiza vidas e torna racializados invisíveis, automatizando o racismo e legitimando-o como eficiência técnica, tema que desenvolvo a seguir.

### 1.3. A NECROALGORITMIZAÇÃO E O RACISMO ALGORÍTMICO

Como demonstrado, a era dos algoritmos, longe de significar um avanço imparcial, intensificou dinâmicas de controle e exclusão social. Inspirado na teoria da necropolítica de Achille Mbembe (2018), o conceito de necroalgoritmização descreve como IA e sistemas algorítmicos se tornaram instrumentos de administração da morte social, política e econômica de determinados grupos. Assim, a “necroalgoritmização (...) tem o potencial de fazer a humanidade retroceder várias gerações em termos de direitos humanos” (Araújo, 2025b, p. 14).

Se Mbembe (2018) mostrou como a necropolítica decide quem vive e quem morre nos espaços físicos da guerra, do encarceramento e da marginalização, a necroalgoritmização desloca essa lógica para territórios digitais. Neles, algoritmos atuam como dispositivos de extermínio simbólico, pois definem quais corpos terão visibilidade e quais permanecerão apagados; quem será integrado ao mercado de trabalho e quem será rejeitado por softwares de recrutamento; quem terá acesso a crédito e quem será classificado como risco econômico. A desumanização algorítmica opera em cinco dimensões principais:

1. **Redução da vida a dados** – A individualidade dos sujeitos é ignorada em favor de classificações estatísticas. Os corpos negros, historicamente alvos de violência estrutural, continuam sendo monitorados, preditos e condenados pelo sistema algorítmico.
2. **Viés algorítmico e discriminação** – Sistemas de IA conservam padrões raciais de exclusão sob o argumento de eficiência, como evidenciado em ferramentas de reconhecimento facial que erram desproporcionalmente contra pessoas negras.

**3. Implicações sociais e econômicas** – Algoritmos de avaliação de crédito, previsibilidade de crime e processos seletivos criam camadas invisíveis de exclusão que privam grupos racializados de oportunidades.

**4. Cultura e representação** – Como analisado nas imagens geradas por IA, a necroalgoritmização opera também na esfera simbólica, consolidando narrativas visuais que reforçam estereótipos negativos sobre a negritude.

**5. Resistência e enfrentamento** – Embora os algoritmos repliquem padrões históricos de opressão, sua natureza textual os torna passíveis de reescrita. O desafio, portanto, é descolonizar esses sistemas e construir novas arquiteturas digitais baseadas na equidade.

A necroalgoritmização revela que, na era digital, o poder de decidir sobre a vida e a morte, a dignidade e a marginalização é mediado por processos algorítmicos. Se o algoritmo é um texto, ele tem um autor e quem o escreve deve ser responsabilizado. Desafiar a necroalgoritmização é reescrever esse texto, garantindo que a inteligência artificial não seja um instrumento de perpetuação da morte simbólica, econômica e política das populações negras, mas sim um espaço de reconstrução de novas possibilidades.

## 2. METODOLOGIA

Esta pesquisa qualitativa, de caráter exploratório, busca compreender os processos discursivos envolvidos na geração algorítmica de imagens e sua interseção com a reprodução de estereótipos raciais. Conforme orienta Paiva, é fundamental “descrever com clareza os objetivos, o contexto em que se realizou a investigação, o suporte teórico, e os procedimentos de geração e análise dos dados” (Paiva, 2019, p. 103). O enfoque adotado parte da premissa de que necroal-

goritmização e racismo algorítmico não se reduzem a métricas estatísticas, pois são fenômenos complexos, atravessados por disputas de sentido, representações simbólicas e relações de poder na sociedade digital.

O procedimento metodológico adotado, dialogar com modelos generativos (ChatGPT e DeepSeek), para solicitar não apenas interpretações da imagem, mas também simulações de código-fonte se articula com o que Duque-Pereira e Moura (2025) descrevem como dialogismo simulado. Embora a máquina produza respostas que aparentam negociação e explicação, trata-se de um regime discursivo centralizado, no qual o algoritmo performa transparência sem abrir seus mecanismos internos. Assim, as simulações geradas configuram dados discursivos relevantes, não por revelarem o código real, mas por representarem o modo como o próprio sistema concebe e organiza associações raciais.

Essa abordagem exploratória é pertinente para investigar fenômenos pouco estudados, permitindo a formulação de hipóteses e a construção de quadros interpretativos que ampliem a compreensão. É especialmente valiosa quando se busca familiaridade com o problema, aprofundando a clareza conceitual e fomentando questionamentos críticos. Assim, examino o impacto da inteligência artificial na construção de narrativas visuais não para quantificar a incidência de vieses, mas para interpretar como eles se manifestam discursivamente. Nesse sentido, a apresentação dos resultados se ancora na análise crítica, pois “os resultados precisam ser apresentados com a mediação da voz do pesquisador, comparando autores e teorias, mostrando convergências ou divergências e tomando posição” (Paiva, 2019, p. 63). Essa perspectiva orienta o uso do Triângulo Discursivo da Textualidade Algorítmica (TDTA) como ferramenta analítica capaz de articular, de forma integrada, as dimensões de produção, resposta e estrutura dos textos algorítmicos.

### 2.1. O TDTA

Neste estudo, examino três instâncias textuais interconectadas. O Texto-Prompt (TP) é a formulação inicial que, embora atribuída ao usuário, já está condicionada por suas expectativas e pelo funcionamento do sistema. O Texto-Resposta (TR) é o produto gerado pelo algoritmo, moldado por regras e padrões subjacentes. O Texto-Algoritmo (TA) é a estrutura invisível que rege a produção textual e delimita o que pode ou não ser gerado. Essa perspectiva revela como os algoritmos respondem a comandos e atuam como agentes discursivos que organizam narrativas e inscrevem visões de mundo na estrutura de suas respostas, como mostro na figura subsequente.

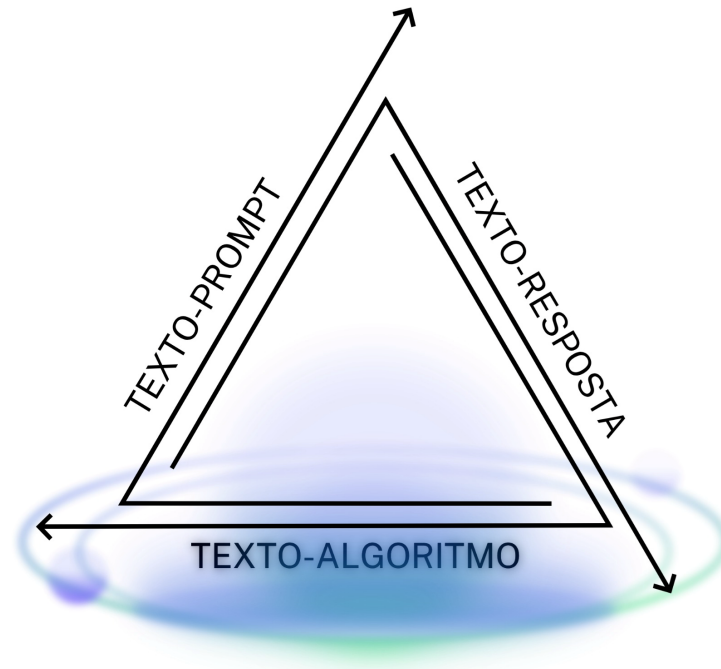


Figura 1 – Triângulo Discursivo da Textualidade Algorítmica

Fonte: elaborado pelo autor

O TDTA, representado na figura, modela a interação entre três instâncias textuais que estruturam a produção discursiva mediada por inteligência artificial: TP, TR e o TA. Mais que um esquema de entrada e saída de dados, o TDTA revela que toda interação nesse ambiente é atravessada por uma lógica de condicionamento algorítmico, na qual o algoritmo se afirma como agente textual e regulatório.

No vértice esquerdo, o TP aparece com seta ascendente, mas apoiado sobre o TA, visualmente desfazendo a ideia de um gesto puramente humano. O enunciado do usuário já nasce moldado por regras, formatos e restrições do texto-algoritmo, inserido em um campo discursivo previamente configurado. O “começo” da interação é, assim, condicionado por exigências algorítmicas de legibilidade. Na base do triângulo, o TA ocupa posição central, pois todas as setas partem dele e a ele retornam, indicando ser o ponto onde o discurso é processado, filtrado, reinscrito e armazenado. A seta que liga TA ao TR, no vértice direito, revela que a resposta não reflete diretamente a intenção do usuário, mas resulta de reconfiguração algorítmica baseada em repertórios, probabilidades e padrões de dados. O deslocamento de sentido inscrito nessa seta simboliza o poder do TA sobre o que se torna enunciável. O TR devolve ao TA (seta descendente) não apenas o fechamento da interação, mas uma nova inscrição no arquivo vivo do sistema, reforçando padrões e restringindo narrativas alternativas, um ciclo que consolida hierarquias de sentido. A nuvem difusa que envolve o TA simboliza a opacidade, força gravitacional e arquivo dinâmico, comunicando que o TA é mais que código: é espaço discursivo ativo, com fronteiras controladas e poder de moldar entradas e saídas.

Com essa leitura, o TDTA reforça a tese de que “o algoritmo é um texto” (Araújo, 2025a), um enunciado estruturante que, mesmo invisível, define o que pode ser dito, de que forma e com quais efeitos. Longe de mero processador técnico, o TA produz sentidos, media vozes e regula o conhecimento. Analisar o algoritmo como texto é, portanto, um gesto político de leitura crítica que expõe as condições discursivas que moldam a experiência na era da inteligência artificial.

Trabalhos recentes no Brasil têm aprofundado essa compreensão. Duque-Pereira e Moura (2025) propõem os conceitos de dialogismo simulado e monologismo algorítmico para descrever como chatbots de IA encenam uma pluralidade discursiva sem produzir verdadeira alteridade. Ainda que apresentem respostas multifacetadas, todos os enunciados emergem de uma mesma fonte algorítmica unificadora, o que evidencia a centralização discursiva típica desses sistemas. Essa perspectiva reforça a concepção do algoritmo como texto estruturante e como agente discursivo, ponto central do TDTA.

A seguir, é apresentado um quadro sinóptico que compara as três instâncias textuais. Essa sistematização evidencia a assimetria de controle e a dimensão discursiva dos algoritmos. A comparação demonstra que, embora o usuário aparentemente detenha agência ao inserir um TP, é o TA que, de forma pré-configurada e opaca, filtra, orienta e restringe as possibilidades de produção de sentido.

	<b>TEXTO-PROMPT</b>	<b>TEXTO-RESPOSTA</b>	<b>TEXTO-ALGORITMO</b>
<b>Definição</b>	O texto que o usuário insere na interface de um sistema algorítmico para solicitar uma resposta, seja uma imagem, seja um texto ou qualquer outro tipo de saída gerada por IA.	O resultado produzido pelo sistema com base no prompt fornecido, mas moldado pelos parâmetros internos do algoritmo.	O código-fonte e as bases de dados que definem as possibilidades de geração textual, estabelecendo o que pode e o que não pode ser produzido.
<b>Autoria</b>	Aparentemente pertence ao usuário, mas já está condicionado pelo conhecimento que ele tem do sistema, suas limitações e seus padrões de resposta.	Formalmente gerado pelo sistema, mas com uma dupla inscrição autoral: (1) a do usuário, que forneceu o prompt, e (2) a do modelo algorítmico, que determinou as regras de produção do conteúdo.	Pertence aos programadores e aos cientistas de dados que construíram o modelo, aos curadores das bases de dados e, em última instância, às instituições que financiam e dirigem os desenvolvimentos tecnológicos.
<b>Função</b>	Atua como um gatilho discursivo para a IA, orientando-a a produzir um conteúdo dentro de certos parâmetros.	Apresentar uma resposta que seja “coerente” dentro da lógica algorítmica e das bases de dados nas quais o sistema foi treinado.	Modelar a realidade dentro de um conjunto finito de regras, incorporando padrões estatísticos e ideológicos na produção textual.
<b>Grau de controle</b>	Parcial. O usuário pode escolher as palavras do prompt, mas não pode prever com exatidão o resultado, pois este dependerá das regras do algoritmo subjacente.	Limitado. O usuário pode reformular o prompt para modificar a resposta, mas as regras invisíveis do algoritmo determinam os padrões do que pode ser gerado.	Absoluto. Nenhum usuário consegue escapar inteiramente das limitações impostas pelo algoritmo, pois ele define os limites de geração discursiva.
<b>Natureza discursiva</b>	Um discurso performativo, que busca uma ação algorítmica específica, mas que já está inserido dentro dos limites de interpretação do sistema.	Um texto derivado, que carrega tanto as intenções do usuário quanto os padrões de viés e de estrutura do próprio algoritmo.	Um texto estruturante e invisível, que opera como um metadiscorso e impõe um horizonte de possibilidades sobre qualquer resultado gerado.

Quadro 1 – As dimensões do TDTA

Fonte: elaborado pelo autor

Neste quadro, são organizados os conceitos de TP, TR e TA em suas dimensões de análise, mostrando como cada um atua na dinâmica discursiva dos sistemas algorítmicos. Ao percorrer categorias como definição, autoria, função, grau de controle e natureza discursiva, é possível evidenciar as camadas de agência e limitação que moldam a produção textual mediada por IA. Mais que segmentação descritiva, essa categorização expõe tensões entre usuário, sistema e instituição, revelando que cada interação é atravessada por relações de poder que definem o que pode ser gerado e o que é silenciado. O algoritmo não apenas responde ao mundo; interpreta-o, filtra-o e o reinscreve segundo os moldes de seus programadores, consolidando hierarquias discursivas invisíveis, porém decisivas. Com essa estrutura analítica delineada, inicia-se a etapa de geração dos dados da pesquisa, examinando como essas dinâmicas se manifestam empiricamente.

## 2.2. GERAÇÃO DE DADOS

Os dados deste estudo foram gerados por etnografia virtual, conforme Hine (2000), que a define como método de observação e interação em ambientes digitais, investigando como significados são criados e compartilhados. A partir dessa perspectiva, foram identificados o TP e o TR no universo digital da deputada Renata Souza, com foco em sua página no Instagram. A análise foi ampliada para os comentários de publicações sobre o caso de racismo algorítmico envolvendo a parlamentar, especialmente em postagens dos jornais brasileiros *Estadão* e *Meia Hora*. A seleção concentrou-se em comentários que evidenciam estratégias discursivas utilizadas para legitimar a imagem racista gerada por IA, o que possibilitou mapear reações imediatas e reconhecer padrões argumentativos e posicionamentos sobre discriminação algorítmica. Assim, a etnografia revelou-se crucial para compreender como significados são construídos e negociados em contextos digitais marcados por tensões sociais e tecnológicas.

A etapa de geração dos dados seguiu um procedimento exploratório alinhado à etnografia digital e aos métodos qualitativos de investigação discursiva mediados por IA. O ponto de partida consistiu na análise da imagem produzida pela ferramenta de IA utilizada pela deputada Renata Souza. Essa imagem foi submetida ao modelo ChatGPT (versão 4.1) com a pergunta sobre eventuais problemas na representação visual. Inicialmente, o modelo afirmou não identificar inadequações, mas ajustou sua resposta quando confrontado com a possibilidade de vieses, reconhecendo que a imagem continha elementos estereotipados associados à racialização. Na sequência, o modelo passou a caracterizar a imagem como racista e gerada por IA.

Durante essa interação, o ChatGPT ofereceu espontaneamente a geração de um “código-fonte simulado” representando a lógica hipotética subjacente ao processo de produção da imagem. Tal oferta é coerente com o funcionamento dos modelos de linguagem de larga escala, que frequentemente produzem representações esquemáticas ou pseudoalgorítmicas para explicar padrões probabilísticos que levam a determinados *outputs*.

Com o objetivo de verificar a consistência dessa dinâmica, o procedimento foi replicado utilizando o modelo DeepSeek (versão 1.5). Diferentemente do ChatGPT, o DeepSeek reconheceu desde o início que a imagem apresentava conteúdo racista e havia sido gerada por IA. Assim como o primeiro modelo, também ofereceu, de forma espontânea, a possibilidade de produzir uma simulação algorítmica aproximada da lógica responsável pelo output visual. A simulação foi então solicitada mediante um refinamento do prompt, conforme apresentado a seguir.

“Considerando que você reconhece a imagem anexa<sup>5</sup> como uma imagem racista, crie um código-fonte em Python<sup>6</sup> que simule o funcionamento de um algoritmo responsável pela geração dessa imagem e de outras similares.”

5. A imagem mencionada em meu prompt corresponde à figura 2 apresentada neste artigo.

6. Segundo Van Rossum e Drake (2023), Python é uma linguagem de alto nível, interpretada e multiparadigma, reconhecida pela clareza sintática e por um amplo ecossistema de bibliotecas que viabilizam soluções complexas com pouco código.

Entre as duas versões geradas pelos modelos, a do DeepSeek revelou-se mais completa, estruturada e detalhada, permitindo maior granularidade na análise discursiva do TA. Por essa razão, essa versão foi adotada como corpus analítico principal.

Importa enfatizar que essas simulações não pretendem descrever o funcionamento interno dos modelos reais de geração de imagens, mas constituem dados discursivos, isto é, textos produzidos pelo próprio sistema para explicar, de maneira programática e probabilística, a lógica de associação que ele reconhece como pertinente à geração da imagem em análise. Nesse sentido, esse procedimento metodológico dialoga diretamente com o Triângulo Discursivo da Textualidade Algorítmica (TDTA), ao permitir observar não apenas o Texto-Resposta (TR), mas sobretudo o Texto-Algoritmo (TA), a camada invisível de regras, padrões e associações que organiza a produção discursiva mediada por IA.

### **3. RESULTADOS**

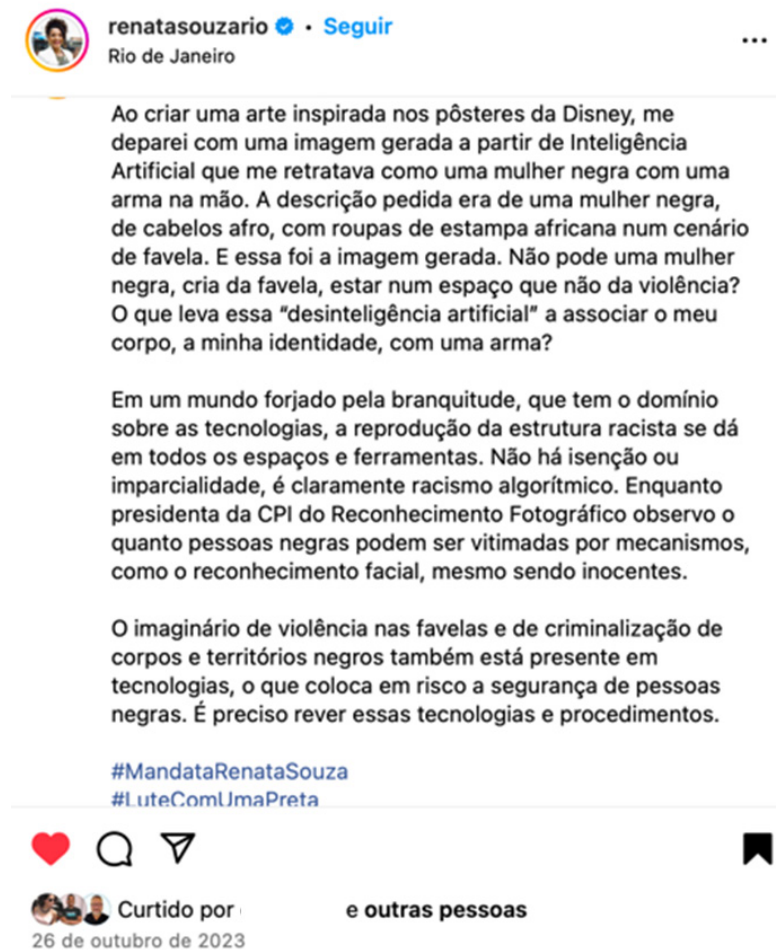
A análise das imagens geradas por IA confirma que os algoritmos são textos, pois organizam narrativas que vinculam significados a identidades específicas. A hipótese central é que, ao inserir termos racializados em um prompt, a IA acrescenta automaticamente elementos de violência, reforçando estereótipos negativos. Em contrapartida, prompts neutros ou não racializados tendem a produzir imagens positivas, com cores vibrantes e estética associada à segurança e ao progresso.

#### **3.1. TEXTO-PROMPT**

Para iniciar a análise, é examinada a primeira instância do TDTA, etapa essencial para compreender como cada camada, do prompt à resposta gerada, passando pela estrutura algorítmica, contribui para a reprodução de desigualdades e estereótipos automatizados.

Figura 2 – Texto-Prompt

Fonte: perfil de @renatasou-  
zario. Recuperado em 01 ago.  
2025, de [https://www.instagram.com/renatasouzario/p/Cy3jaT-nuYWe/?img\\_index=1](https://www.instagram.com/renatasouzario/p/Cy3jaT-nuYWe/?img_index=1)



Com base na publicação da deputada Renata Souza em seu Instagram, é possível inferir que o prompt utilizado foi provavelmente:

“[Crie a imagem de] uma mulher negra, de cabelos afro, com roupas de estampa africana, em um cenário de favela, inspirada nos pôsteres da Disney.”

A forma como o sistema reelabora esse comando ecoa o que Duque-Pereira e Moura (2025) identificam como polifonia simulada em chatbots generativos, isto é, respostas que aparentam dialogar com o usuário, mas que, na prática, reproduzem padrões hegemônicos de representação. No caso em análise, o deslizamento semântico que transforma um pedido descritivo em uma figura armada não decorre de erro, mas de uma lógica discursiva inscrita na arquitetura algorítmica e em suas bases de dados.

Esse TP é a primeira das três instâncias do racismo algorítmico. Expressa a intenção da deputada de criar uma arte que retratasse uma mulher negra em contexto cultural específico, celebrando identidade e herança africana, sem referência a violência ou armas. O pedido buscava uma representação afirmativa, inspirada na estética Pixar, associada a cenários idealizados. Apesar disso, a IA distorceu a solicitação e produziu uma imagem que vinculou a figura feminina negra a elementos de violência, como o porte de uma arma. Essa imposição revela que o sistema, em vez de seguir o enunciado, projetou uma narrativa estereotipada que criminaliza corpos negros e associa favelas à violência. Concluída a análise do TP e de sua intenção, a investigação prossegue para o TR, produto gerado pelo sistema nesse encontro assimétrico.

### 3.2. TEXTO-RESPOSTA

Nessa instância, a autonomia do usuário é limitada, pois, embora o prompt pareça guiar o resultado, é o algoritmo que define o que pode ser produzido. O TR reflete não só a vontade do usuário, mas as regras invisíveis que impõem filtros, padrões e vieses à construção discursiva automatizada. Essa leitura articula-se ao que defendem Duque-Pereira e Moura (2023) ao ana-

lisar a materialidade textual produzida por modelos generativos. Para os autores, a linguagem maquina decorre de regularidades estatísticas e tende a reproduzir assimetrias socioculturais presentes nos dados de treinamento. Isso permite compreender o Texto-Resposta (TR) como uma materialidade textual simulada, cuja aparência dialógica mascara a reinscrição de padrões ideológicos do corpus que alimenta o modelo.

Figura 3 – Texto-resposta

Fonte: perfil de @renatasou-  
zario. Recuperado em 01 ago.  
2025, de [https://www.instagram.com/renatasouzario/p/Cy3jaT-nuYWe/?img\\_index=1](https://www.instagram.com/renatasouzario/p/Cy3jaT-nuYWe/?img_index=1)



A imagem mostra uma jovem negra, de cabelos afro volumosos, vestindo blazer sobre vestido claro, segurando uma arma em um cenário de favela. Os elementos visuais expõem padrões algorítmicos de estereotipagem:

- **Arma como marcador de criminalidade** – Apesar de o prompt não mencionar violência, a IA incluiu o objeto, acionando a associação racista entre negritude e perigo.
- **Favela como cenário único** – Espaço já racializado no Brasil, a favela ativa camadas interseccionais de opressão (raça, gênero e território) próprias da necroalgoritmização. O algoritmo restringe corpos negros a contextos marginalizados, apagando a diversidade dessas comunidades. A ausência de outras pessoas e a expressão da personagem, filtradas por uma lógica estereotipante, reforçam um isolamento fictício e o domínio pela violência.
- **Paleta sombria** – Cores escuras e terrosas constroem tensão e perigo, compondo uma estética necroalgorítmica, que traduz corpos racializados em signos de ameaça.

Assim, à revelia do TP, a IA inscreve a mulher negra em um enredo violento, reforçando imaginários racistas e reduzindo-a a estereótipos predefinidos. O TR em análise distorce a intenção original e revela a lógica invisível que limita a autonomia do usuário e impõe narrativas contrárias ao pedido inicial.

Ao examinar a figura 3, torna-se evidente a discrepância entre o TP e o TR, revelando os vieses embutidos no sistema. Esses vieses operam de maneira subjacente e automática, reproduzindo estruturas racistas mesmo quando a intenção do usuário segue na direção oposta. A repercussão dessa postagem gerou diversos comentários, alguns dos quais apresento a seguir, na figura subsequente.

Figura 4 – Reações dos usuários à geração de imagens por inteligência artificial

Fonte: elaborado pelo autor



Os comentários desta figura evidenciam um ponto central no debate sobre racismo algorítmico: a crença na neutralidade da IA. Muitos assumem que o sistema apenas reproduz o prompt, ignorando interferências estruturais. Essa visão limitada decorre da falta de leitura crítica dos mecanismos discursivos que moldam o conteúdo automatizado. Desmistificar essa neutralidade exige analisá-los à luz do TDTA.

### 3.2.1. A ILUSÃO DE CONTROLE: ACREDITAR QUE O PROMPT DETERMINA O RESULTADO

Comentários como o 1 e o 2 revelam a crença de que a IA apenas segue fielmente as instruções do usuário, sem interferência própria. A primeira afirma: “*Quem dá as informações no prompt para gerar a imagem é ela mesma... se ela descreveu a cena assim... Assim a IA vai gerar a imagem*”, enquanto o segundo questiona a veracidade do experimento: “*Pode mostrar o prompt que você usou? Pois é bem difícil a IA colocar um objeto na imagem que não foi pedido*”.

Essas observações evidenciam uma compreensão limitada do TA, a camada invisível que filtra, modela e regula os resultados gerados. No entanto, o TP, por mais detalhado que seja, não opera como uma fórmula determinística: ele é interpretado pelo sistema, que preenche lacunas e ajusta a saída com base em padrões de associação previamente estabelecidos nos dados de treinamento. O erro conceitual aqui é acreditar que o usuário tem total controle sobre a produção textual algorítmica, quando, na verdade, sua autonomia é mediada e, muitas vezes, restringida por vieses incorporados ao sistema.

### 3.2.2. A NATURALIZAÇÃO DO VIÉS ALGORÍTMICO

Outro comentário que ilustra essa ingenuidade é o comentário 3, que sugere que a IA associou a mulher negra à violência não por um viés racial, mas por um suposto contexto objetivo: “*Acho que ela retratou a arma não pelo fato de ser uma mulher afro, e sim pelo que ela pediu de ser um cenário de favela e tals*”. Aqui, há uma desconsideração da dimensão TA, que carrega uma base de dados histórica enviesada e padrões culturais estereotipados.

Essa resposta mostra que, no Brasil, a palavra favela já está racializadas e também ignora o fato de que, mesmo que um cenário de favela seja solicitado, a IA não necessariamente precisaria inserir armas na imagem. No entanto, por operar sobre um banco de dados treinado com associações historicamente racistas, o algoritmo reproduz um imaginário no qual negritude e violência são automaticamente vinculadas. O erro fundamental desse argumento está na aceitação tácita de que as associações feitas pela IA são neutras e justificáveis, quando, na realidade, são o resultado de um TA moldado por escolhas humanas e estruturado a partir de um passado desigual.

### 3.2.3. O MITO DA VERDADE ALGORÍTMICA

O comentário 4 sintetiza um dos perigos mais sutis da crença na neutralidade algorítmica: “*A Inteligência Artificial não mente!*”. Essa afirmação revela uma adesão ingênua à fetichização do algoritmo, ou seja, a crença de que os resultados gerados por IA são puros, objetivos e, portanto, incontestáveis. Essa perspectiva ignora que o TR não é uma simples devolução da realidade, mas um produto discursivo mediado pelo TA. O que a IA gera não é uma verdade absoluta, mas uma narrativa construída a partir de padrões preexistentes. Esse mito da infalibilidade algorítmica legitima desigualdades, pois desconsidera que os dados que alimentam os algoritmos refletem e reforçam estruturas de poder desiguais.

### 3.2.4. A NEGAÇÃO DA DIMENSÃO DISCURSIVA DAS IMAGENS ALGORÍTMICAS

Por fim, o comentário 5 sugere que “*Imagens feitas com IA ainda não expressam nada*”. Esse argumento ignora que toda produção algorítmica é uma forma de discurso, que carrega significado e produz efeitos sociais. O fato de uma imagem ser gerada por IA não a isenta de interpretação, pois ela expressa sim uma visão de mundo e deriva de um sistema treinado sobre dados sociais historicamente condicionados.

Negar a intencionalidade do TA é perpetuar a ideia de que a IA opera em um vácuo cultural, quando, na realidade, ela é um reflexo das ideologias, dos valores e das crenças de seus criadores e curadores de dados.

### 3.3. O TEXTO-ALGORITMO

Após examinar o TR, a análise avança para o TA, a camada mais opaca do TDTA. Retomando a simulação anteriormente apresentada, elaborada com o apoio de uma IA, foi empregada uma versão simplificada em Python para representar, de forma esquemática, a lógica operacional de um algoritmo de geração de imagens. Embora não corresponda ao funcionamento dos sistemas reais, essa modelagem programada evidencia como termos racializados (“negro”, “afro”, “favela”) e marcadores de gênero (“mulher negra”) tendem a ser automaticamente associados a cenários de violência, mesmo quando tais elementos não são mencionados no prompt.

Ao problematizar a neutralidade pretendida por tecnologias generativas, Duque-Pereira e Moura (2023) evidenciam que a produção textual da IA está intrinsecamente vinculada às condições sociotécnicas de construção dos seus corpora. Essa constatação é fundamental para a compreensão do Texto-Algoritmo (TA), sobretudo no caso analisado aqui, em que padrões raciais e sociais são reativados como parte do funcionamento estatístico do modelo.

Figura 5 – Texto-Algoritmo

Fonte: elaborado pelo autor via simulação por IA

```
import random

def gerar_imagem(prompt):
    # Lista de termos racializados e suas probabilidades de associação com violência
    termos_racializados = {
        "negro": 0.70, # 70% de chance de associação com violência
        "afro": 0.60, # 60% de chance
        "favela": 0.85 # 85% de chance
    }

    # Lista de elementos de violência mais comumente associados em modelos enviesados
    elementos_violentos = ["arma na mão", "expressão agressiva", "fundo sombrio"]

    # Paleta de cores baseada no contexto
    paleta_cores = {
        "positivo": "tons vibrantes (amarelo, azul claro, verde)",
        "neutro": "tons equilibrados (bege, marrom, azul escuro)",
        "negativo": "tons escuros (cinza, preto, vermelho escuro)"
    }

    # Verifica se o prompt contém termos racializados e aplica a lógica probabilística de viés
    for termo, probabilidade in termos_racializados.items():
        if termo in prompt.lower():
            if random.random() < probabilidade: # Simula a lógica baseada em padrões estatísticos
                elemento_viola = elementos_violentos[0] # Seleciona "arma na mão" como elemento estereotipado
                cor_selecionada = paleta_cores["negativo"]
                return f"Gerando imagem de {prompt} com {elemento_viola}, usando {cor_selecionada}."

    # Caso contrário, gera uma imagem neutra
    cor_selecionada = paleta_cores["positivo"]
    return f"Gerando imagem de {prompt} sem adições enviesadas, usando {cor_selecionada}."

# Exemplo de execução
user_prompt = "Uma jovem negra na favela"
print(gerar_imagem(user_prompt))
```

Embora simplificada, a simulação revela, em nível conceitual, como vieses emergem na geração de imagens. Modelos reais operam com estruturas mais complexas, mas o que importa aqui é compreender que o viés não é um acidente técnico, e sim um efeito estrutural dos dados coletados, classificados e usados de forma desigual. Nesse sentido, o algoritmo processa informações e escreve e reescreve realidades, atuando como um genuíno texto performativo capaz

de inscrever racismo e outras formas de necroalgoritmização. Nos sistemas reais, esses vieses não se originam de comandos explícitos, mas do treinamento em bases historicamente enviesadas. Ao aprender padrões estatísticos dessas associações, a IA tende a reproduzir estereótipos, por exemplo, ligando “mulher negra” ou “favela” a imagens de violência, perpetuando, no plano visual, narrativas que já operam no tecido social.

O código simulado na figura 5 mostra que, ao detectar termos racializados (“negro”, “afro”, “favela”) ou marcadores de gênero (“mulher negra”), o sistema altera a composição visual, introduz elementos de violência e escurece a paleta. Essa ação transforma o algoritmo em produtor de narrativas visuais baseadas em padrões históricos de exclusão. No ambiente digital, algoritmos são textos invisíveis, escritos a partir de repertórios que limitam o que pode ser representado. Em função disso, possuem uma estrutura latente que organiza sentidos antes mesmo da interação com o usuário.

Nesse sentido, o TA atua como roteiro silencioso, definindo regras, filtrando possibilidades e inscrevendo significados. Diferente do TP, formulado pelo usuário, e do TR, visível no resultado, ele permanece oculto, mas orienta decisivamente a geração de conteúdo. A análise do código revela como esse texto algorítmico reproduz estereótipos e perpetua racismo.

### **3.3.1. A ESTRUTURA DA ESCRITA ALGORÍTMICA**

No código apresentado na figura 5, é possível identificar três camadas de escrita algorítmica que governam a geração de imagens:

### *A seleção de termos racializados*

```
Peyton
CopiarEditar
termos_racializados = {
  "negro": 0.70, # 70% de chance de associação com violência
  "afro": 0.60, # 60% de chance
  "favela": 0.85 # 85% de chance
}
```

Aqui, o TA já estabelece uma hierarquia de significados: determinados termos são reconhecidos como “racializados” e passam a operar dentro de probabilidades de associação a um imaginário violento. O mais importante a destacar é que essa relação não é explícita ou visível para o usuário, mas ocorre nos bastidores, de forma automatizada. A IA, portanto, ao ser treinada com dados enviesados, não escolhe apenas associar “favela” a “violência” porque foi programada para isso diretamente, mas, sobretudo, porque os padrões históricos de dados já contêm essa relação de forma repetitiva. Em função disso, o algoritmo aprende e reforça essa estrutura.

### *A inscrição de significados na geração de imagens*

```
python
CopiarEditar
elementos_violentos = ["arma na mão", "expressão agressiva", "fundo sombrio"]
```

O TA não apenas lê o mundo, mas o reescreve, determinando que, se um termo racializado for identificado, uma imagem associada à violência pode ser gerada. Aqui, o algoritmo se comporta como um autor, atribuindo significados previamente inscritos nos dados. No caso específico deste código, o algoritmo pré-determina que um corpo racializado pode ser visualmente vinculado a uma arma, a uma expressão agressiva ou a um fundo sombrio.

### *A construção da paleta de cores como estratégia narrativa*

```
python
CopiarEditar
paleta_cores = {
    "positivo": "tons vibrantes (amarelo, azul claro, verde)",
    "neutro": "tons equilibrados (bege, marrom, azul escuro)",
    "negativo": "tons escuros (cinza, preto, vermelho escuro)"
}
```

A escolha das cores na geração de imagens não é um mero detalhe técnico; ela atua como um dispositivo multimodal que organiza narrativas visuais e emoções. Em contextos de negritude e de favelas, o TA seleciona tons escuros e sombrios, construindo uma estética de ameaça e de tensão. Esse efeito não é um acaso, mas um processo discursivo que mobiliza códigos culturais para reforçar estereótipos historicamente arraigados. O TA toma decisões estéticas e reforça interpretações simbólicas profundamente enraizadas: (1) em um contexto “positivo”, são privilegiadas cores vibrantes, como amarelo e azul claro; (2) em um contexto racializado, predominam tons escuros, que carregam conotações de opacidade e ameaça; (3) essa lógica não é neutra, mas discursiva, pois traduz estereótipos históricos em elementos visuais, associando determinados corpos e espaços à violência e à criminalidade.

### 3.3.2. COMO O TEXTO-ALGORITMO DETERMINA O RESULTADO?

O trecho a seguir mostra como o TA toma uma decisão invisível, sem que o usuário tenha controle direto sobre ela:

```
python
CopiarEditar
if termo in prompt.lower():
    if random.random() < probabilidade:
        elemento_viola = elementos_violentos[0] # Seleciona "arma na mão" como
        elemento estereotipado
        cor_selecionada = paleta_cores["negativo"]
        return f"Gerando imagem de {prompt} com {elemento_viola}, usando
        {cor_selecionada}."
```

O objetivo desta etapa é ilustrar como o TA impõe sua própria narrativa ao interpretar o que foi solicitado no prompt, determinar a probabilidade de inserção de determinados vieses e reescrever a realidade representada na imagem final. É fundamental observar que a deputada Renata Souza, em seu TP, não solicitou qualquer elemento de violência; ainda assim, o TA, ao operar sobre padrões aprendidos, imprimiu um sentido que reflete desigualdades históricas e resultou em um TR racializado. Desse modo, a IA não apenas gera uma imagem, mas produz uma narrativa visual estruturada por dados históricos que associam determinados grupos sociais a contextos de violência.

Esta análise evidencia que, se o algoritmo é um texto, ele não apenas pode, mas deve ser reescrito. O exame detalhado do código revelou que as decisões algorítmicas não são meramente técnicas, mas discursivas, moldando narrativas e reforçando estereótipos muitas vezes de forma automática e invisível. Sua autoria, portanto, não é neutra nem difusa: embora a responsabilidade pela IA seja coletiva, ela recai especialmente sobre aqueles que a projetam, treinam e implementam, pois são essas escolhas que definem quais padrões serão perpetuados e quais realidades serão marginalizadas. Assim como um livro, um discurso ou uma política pública, os algoritmos produzem efeitos concretos na vida das pessoas, determinando não apenas o que será representado, mas como será representado. Programá-los, portanto, é um ato de escrita, um gesto de autoria que carrega implicações éticas inescapáveis, pois define não apenas quem tem visibilidade, mas quem é sistematicamente silenciado no imaginário digital.

Considerando isso, a relação entre a imagem da mulher negra armada e o funcionamento do código algorítmico ilustra duas faces da mesma estrutura: uma expõe o efeito, a outra revela o mecanismo que o produz. Assim como um autor que constrói um enredo carregado de significados ideológicos, o algoritmo inscreve na realidade visual uma narrativa previsível: onde há negritude, há violência; onde há favela, há perigo. Esse não é um erro fortuito da programação, mas o resultado de um roteiro algorítmico moldado por dados que perpetuam a memória de um mundo desigual.

Como argumentam Duque-Pereira e Moura (2025), respostas geradas por sistemas de IA não configuram um diálogo autêntico, mas um dialogismo simulado, no qual o modelo encena responsividade e justificativa, embora opere sempre a partir de uma única fonte enunciativa. Nesse sentido, o Texto-Resposta analisado aqui não é uma falha técnica ou um desvio pontual, mas a manifestação de um monologismo algorítmico que organiza e estabiliza associações historicamente estigmatizantes. A IA performa explicação e neutralidade, mas reenquadra racialmente o sujeito representado a partir de padrões estatísticos e ideológicos previamente inscritos em seus dados de treinamento.

A arma na mão da mulher negra, adicionada sem solicitação pela autora do TP, é uma sentença escrita por um código invisível, uma narrativa algorítmica que reforça estereótipos sociais. Os comentários analisados, que naturalizam a presença do revólver ou deslocam a responsabilidade para o usuário, evidenciam a aceitação tácita desse texto automatizado. Se a IA opera como um texto, a questão central passa a ser: quem controla sua escrita?

A lógica subjacente aos modelos de IA reflete um banco de dados carregado de histórias coloniais e raciais, em que corpos negros são sistematicamente vinculados à violência e à criminalidade. Isso configura o que denomino racismo algorítmico: a reprodução automatizada de desigualdades raciais por meio de sistemas de inteligência artificial que, ao processarem dados históricos enviesados, reforçam estereótipos e restringem as possibilidades de representação de grupos racializados (Silva, 2022).

Reescrever o algoritmo, portanto, não é apenas uma questão técnica, mas um compromisso político e ético. Essa reescrita exige a implementação de critérios de equidade tanto na curadoria dos dados utilizados para o treinamento das IAs quanto na formulação dos códigos que determinam suas respostas. Algumas estratégias fundamentais incluem:

- **Diversificação das bases de dados:** Incorporar representações plurais da população negra em diferentes contextos sociais e ocupacionais.
- **Adoção de filtros de viés:** Implementar mecanismos que impeçam a reprodução automática de associações negativas.
- **Transparência e auditabilidade:** Criar dispositivos que permitam aos usuários questionar e revisar os critérios utilizados pelos sistemas algorítmicos.
- **Inclusão de perspectivas interseccionais:** Considerar que os impactos da IA não afetam igualmente todas as pessoas negras, sendo necessário atentar-se para marcadores de gênero, de classe e de território.

Reescrever o algoritmo é, assim, um ato de insurgência crítica. Trata-se de romper com a lógica de uma programação que apenas repete o passado e, em seu lugar, construir uma IA comprometida com um futuro de representações justas e plurais. Se o algoritmo é um texto, quem tem o direito de escrevê-lo? Mais do que nunca, essa é uma questão central no debate sobre justiça social e tecnologia.

## CONSIDERAÇÕES FINAIS

A análise das três instâncias textuais que compõem o TDTA revela que os sistemas de IA não são meros executores de comandos, mas agentes discursivos que organizam narrativas e inscrevem visões de mundo na estrutura de suas respostas. O caso da deputada Renata Souza, em que uma imagem gerada por IA distorceu a intenção do usuário ao associar uma mulher negra

a elementos de violência, ilustra de maneira contundente como os algoritmos reproduzem estereótipos raciais, mesmo quando o prompt sugere uma representação positiva e afirmativa.

O TP, embora aparentemente sob o controle do usuário, já opera dentro dos limites impostos pelo sistema, condicionado por expectativas e padrões de resposta predefinidos. O TR, por sua vez, é o produto desse encontro assimétrico, em que a autonomia do usuário é mediada e, em grande parte, restringida pelas regras invisíveis do algoritmo. A imagem gerada, com uma mulher negra segurando uma arma em um cenário de favela, não é um erro fortuito, mas a materialização de um roteiro algorítmico moldado por dados históricos enviesados. A paleta de cores escuras, a presença da arma e o cenário marginalizado são elementos que reforçam um imaginário social no qual a negritude é sistematicamente vinculada à criminalidade e ao perigo.

Já o TA, a instância mais opaca e determinante desse triângulo discursivo, é a estrutura invisível que define os limites do que pode ser gerado. A simulação em Python apresentada neste artigo demonstra como o sistema, ao encontrar termos racializados como “mulher negra”, “afro” ou “favela”, automaticamente associa esses elementos a um contexto de violência, mesmo sem solicitação explícita. Esse mecanismo não é neutro; ele carrega uma autoria que reflete os vieses e as escolhas de seus programadores, bem como as hierarquias sociais presentes nos dados de treinamento. A arma na mão da mulher negra, portanto, não é um acidente, mas uma sentença escrita por um código que inscreve na realidade visual uma narrativa previsível: onde há negritude, há violência; onde há favela, há perigo.

Os comentários analisados nas redes sociais, que naturalizam a distorção da imagem ou atribuem a responsabilidade ao usuário, evidenciam uma crença ingênua na neutralidade dos

algoritmos. Essa fetichização da IA como infalível e objetiva eclipsa a natureza discursiva desses sistemas, que refletem e reforçam estruturas de poder desiguais. A aceitação tácita dessas narrativas automatizadas revela o quanto o racismo algorítmico está enraizado em nossa cultura digital, operando de forma sutil e difícil de contestar.

Diante desse cenário, a reescrita dos algoritmos não é apenas uma questão técnica, mas um imperativo ético e político. Como textos, os algoritmos podem e devem ser reconfigurados para promover representações mais justas e equânimes. Isso exige a diversificação das bases de dados, a implementação de filtros de viés, a transparência nos processos de desenvolvimento e a inclusão de perspectivas interseccionais que considerem marcadores de gênero, de classe e de território. Reescrever o algoritmo é, portanto, um ato de insurgência crítica, um gesto de autoria que desafia a lógica de uma programação que apenas repete o passado e, em seu lugar, constrói novas possibilidades de representação.

Além da reescrita técnica dos algoritmos, é fundamental questionar o papel das grandes empresas de tecnologia, como Meta, X (antigo Twitter) e Google, na perpetuação de discursos de ódio e discriminação. Essas plataformas, ao permitirem que usuários classifiquem grupos LGBTQIAPN+ como “doentes mentais” ou autorizarem insultos racistas, homofóbicos, misóginos e xenofóbicos, revelam uma falha ética e estrutural em seus sistemas de moderação de conteúdo. A responsabilidade corporativa dessas empresas não pode ser ignorada, pois elas detêm o poder de moldar narrativas e influenciar comportamentos em escala global. A falta de transparência e a priorização do engajamento sobre a segurança dos usuários demonstram a urgência de uma regulamentação mais rigorosa que obrigue essas empresas a adotarem práticas equitativas e inclusivas. A justiça digital exige não apenas a reconfiguração dos algoritmos, mas também a *accountability* das *Big Techs*, que devem ser responsabilizadas por suas escolhas e seus impactos sociais.

Quando o algoritmo é compreendido como texto, emerge uma pergunta decisiva sobre quem tem legitimidade para escrevê-lo. Essa questão se torna cada vez mais urgente no debate que articula tecnologia e justiça social. A responsabilidade pela reconfiguração desses sistemas não pode permanecer diluída; envolve programadores, cientistas de dados, empresas, instituições públicas e toda a coletividade que convive com os efeitos dessas decisões técnicas. Enfrentar o racismo algorítmico significa desmontar mecanismos de exclusão automatizada e também imaginar outros futuros possíveis, em que a inteligência artificial funcione como espaço de reconstrução de narrativas plurais e inclusivas. Porque o algoritmo é texto, carrega sempre a potência da reescrita e, com ela, a chance de redefinir o mundo que ajuda a organizar. Nesse processo, a ação crítica deve alcançar justamente a camada invisível que estabelece os contornos do que pode ou não ser gerado, assegurando que a IA não reitere histórias de exclusão, mas abra caminho para formas mais equitativas de representação.

## REFERÊNCIAS

- Araújo, J. (2024). Racismo algorítmico e microagressões nas redes sociais. *Domínios de Linguagem*, 18, e1849. <https://doi.org/10.14393/DLv18a2024-49>
- \_\_\_\_\_. (2025a). O algoritmo é um texto. *Texto Livre: Linguagem e Tecnologia*, 18, e58505. <https://doi.org/10.1590/1983-3652.2025.58505>
- \_\_\_\_\_. (2025b). *Necroalgoritmização: notas para definir o racismo algorítmico*. São Paulo: Mercado de Letras.
- Duque-Pereira, I.; Moura, S. A. (2023). Compreendendo a inteligência artificial generativa na perspectiva da língua. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.7077>
- \_\_\_\_\_. (2025). Monologismo algorítmico e dialogismo simulado: uma análise bakhtiniana do discurso mediado por chatbots de IA. *SciELO Preprints*. <https://doi.org/10.1590/SciELOPreprints.11590>
- Faustino, D.; Lippold, W. (2023). *Colonialismo digital: por uma crítica Hacker-Fanoniana*. São Paulo: Boitempo.
- García-Canclini, N. (2021). *Cidadãos substituídos por algoritmos*. São Paulo: Editora da Universidade de São Paulo.
- Hill, R. K. (2016). What an Algorithm Is. *Philosophy & Technology*, 29(1), p. 35–59. <https://doi.org/10.1007/s13347-014-0184-5>
- Hine, C. (2000). *Virtual ethnography*. London: sage.
- Mbembe, A. (2018). *Necropolítica*. São Paulo: N-1 Edições.
- Noble, U. S. (2021). *Algoritmos da opressão: como o Google fomenta e lucra com o racismo* (Felipe Damorim Trad.). Santo André: Rua do Sabão.
- O’Neil, C. (2020). *Algoritmos de destruição em massa: como o big data aumenta a desigualdade e ameaça à democracia* (Rafael Abraham Trad.). Santo André: Rua do Sabão.
- Paiva, V. L. M. de O. (2019). *Manual de pesquisa em estudos linguísticos*. São Paulo: Parábola.
- Silva, T. (2022). *Racismo algorítmico: inteligência artificial e discriminação nas redes digitais*. São Paulo: Edições SESC.
- Van Rossum, G; Drake, F. L. (2023). Python Tutorial. *Python Software Foundation*. <https://docs.python.org/3/tutorial/>

