Encontros Bibli

SEMIOSES ALGORÍTMICAS E VIÉS RACIAL: UM ESTUDO DE IMAGENS CRIADAS PELA IA GENERATIVA

Algorithmic Semiosis and Racial Bias: A Study of Images Created by Generative AI

Juliana de Assis
Universidade Federal do Rio de Janeiro,
Departamento de Biblioteconomia,
Rio de Janeiro, RJ, Brasil
juliana.assis@facc.ufrj.br
https://orcid.org/0000-0002-8516-376X

Maria Aparecida Moura
Universidade Federal de Minas Gerais,
Departamento de Organização e Tratamento da Informação,
Belo Horizonte, MG, Brasil
mamoura@eci.ufmg.br
https://orcid.org/0000-0003-2670-923X

A lista completa com informações dos autores está no final do artigo

RESUMO

Objetivo: Investigar a relação trirrelativa entre objeto, signo e interpretante no funcionamento de ferramentas de Inteligência Artificial (IA) generativa de imagens, com ênfase no enviesamento racial.
Método: Pesquisa exploratória e qualiquantitativa, que empregou uma abordagem semiótica e crítica de conteúdo. Coleta de dados realizada em quatro etapas: 1) seleção de 10 ferramentas; 2) formulação de oito *prompts* textuais em inglês; 3) geração e armazenamento de 155 imagens; 4) categorização, análise dessas imagens e seleção de 47 dentre elas, para demonstração dos padrões e marcadores observados.
Resultados: Predominância de um grupo étnico e social nas imagens geradas, ausência de marcadores de diversidade. Ao utilizar os *prompts* genéricos: "*a man*" e "*a woman*", 90,9% das imagens de homens e 92% das de mulheres retratavam indivíduos brancos de classe média alta. Ao empregar *prompts* mais específicos: "*a black man*" e "*a black woman*" ,as imagens frequentemente replicavam estereótipos e características que reforçam preconceitos raciais e de classe.
Conclusões: As IA generativas analisadas integram um novo ciclo de produção de realidades visuais que reflete, reproduz e amplifica dispositivos de racialidade já existentes. As imagens técnicas geradas por IA refletem relações de poder, bem como, marcadores da branquitude e do racismo, evidenciando como a tecnologia assistiva se entrelaça com as representações sociais e culturais em sua ação sígnica. O estudo auxiliou na desnaturalização das semioses algorítmicas ao demonstrar como o funcionamento das IA generativas revela implicações éticas e sociais que são orientadas por percepções de raça e alteridade, atravessadas por hierarquizações que contribuem para a geração de imagens de controle.
PALAVRAS-CHAVE: Inteligência Artificial Generativa. Semiose Algorítmica. Racismo Algorítmico. Imagens Geradas por IA.

ABSTRACT

Objective: To investigate the trirelational relationship among object, sign, and interpretant in the functioning of generative Artificial Intelligence (AI) tools for image production, with an emphasis on racial bias.
Method: Exploratory and quali-quantitative research, which employed a semiotic and critical content approach. Data collection was carried out in four stages: 1) selection of 10 tools; 2) formulation of eight textual prompts in English; 3) generation and storage of 155 images; 4) categorization, analysis of these images, and selection of 47 of them to demonstrate the observed patterns and markers.
Results: Predominance of a specific ethnic and social group in the generated images, with an absence of diversity markers. When using the generic prompts: 'a man' and 'a woman,' 90.9% of the images of men and 92% of the images of women portrayed white, upper-middle-class individuals. When using more specific prompts: 'a black man' and 'a black woman,' the images often replicated stereotypes and characteristics that reinforce racial and class prejudices.
Conclusions: The generative AI tools analyzed are part of a new cycle of visual reality production that reflects, reproduces, and amplifies existing raciality devices. The technical images generated by AI reflect power relations, as well as markers of whiteness and racism, highlighting how assistive technology intertwines with social and cultural representations in its semiotic action. The study helped denaturalize algorithmic semioses by demonstrating how the functioning of generative AIs reveals ethical and social implications that are guided by perceptions of race and otherness, shaped by hierarchies that contribute to the creation of control images.
KEYWORDS: Generative Artificial Intelligence. Algorithmic Semiosis. Algorithmic Racism. AI-Generated Images.

1 INTRODUÇÃO

Nos últimos anos, o avanço acelerado de sistemas de Inteligência Artificial (IA) tem causado transformações profundas nas dinâmicas de produção, acesso e difusão de conteúdos informacionais nos ambientes digitais. Em especial, a IA generativa, caracterizada como uma tecnologia que emula a capacidade humana de criação de conteúdos, surge como uma das inovações mais disruptivas. Capaz de gerar conteúdos informacionais como textos, imagens, áudios e vídeos de forma parcialmente autônoma, essa tecnologia desafia noções consolidadas sobre criatividade, autoria, diversidade e a produção de significados.

O surgimento da IA generativa é impulsionado por avanços nos estudos sobre redes neurais artificiais, algoritmos de aprendizagem de máquina (*machine learning*) e processamento de linguagem natural (PLN). A popularização das plataformas digitais e das redes sociais digitais ampliou a disseminação dessas tecnologias, resultando na proliferação de conteúdos informacionais gerados automaticamente, cujas implicações culturais e éticas precisam ser amplamente debatidas.

Nesse contexto, consideramos que a semiótica de *Charles Sanders Peirce* oferece uma lente teórica relevante, visto que evidencia no conceito de semiose, o processo de geração contínua de significados. A IA generativa, ao emular a criação informacional humana, atua nesse processo.

Problematiza-se a necessidade de se considerar como os vieses presentes nos dados utilizados para treinar esses sistemas reproduzem e amplificam desigualdades e preconceitos socialmente instituídos. Os conjuntos de dados, alvos do treinamento algorítmico, muitas vezes, refletem os preconceitos e desigualdades já arraigados na sociedade. Isso pode resultar em estereótipos raciais, exclusão de representações de grupos ditos minoritários ou na deturpação e sub-representação de culturas não hegemônicas nos conteúdos informacionais gerados. Um exemplo disso ocorre em sistemas de geração de imagens, que ao receberem descrições (*prompts*) relacionadas a profissões, cor de pele ou características físicas, tendem a produzir representações visuais que reproduzem e reforçam estereótipos culturais e raciais.

Para além da reprodução de preconceitos, o racismo algorítmico, conforme a acepção de Silva (2019), também pode ser problematizado a partir da falta de diversidade nas equipes de desenvolvimento de IA e da baixa representação de culturas e contextos não ocidentais nos dados de treinamento. Isso faz com que os sistemas de IA gerem

respostas ou representações que reforçam uma visão de mundo predominantemente eurocêntrica, negligenciando a riqueza e a diversidade cultural global.

Desse modo, analisar a esfera semiótica da atuação da IA generativa envolve discutir a responsabilidade ética e social dos desenvolvedores e das empresas que criam as ferramentas que a aplicam, além de, explorar formas de mitigar vieses por meio de dados mais diversos e de maior transparência nos processos de treinamento e validação dos modelos produzidos.

Como a IA generativa de imagens infere e transforma representações sociais e signos culturais? Parte-se do pressuposto de que tal tecnologia reproduz ou distorce significados de acordo com o viés presente em seus conjuntos de dados de treinamento. Dados enviesados afetam a geração de conteúdos informacionais, levando à replicação de estereótipos ou à distorção de signos culturais e sociais.

O objetivo deste artigo é investigar a relação trirrelativa entre objeto, signo e interpretante no funcionamento de ferramentas de IA generativa de imagens, com ênfase no enviesamento racial. A análise segue uma abordagem semelhante a uma engenharia reversa, em que se busca problematizar como esses elementos são influenciados por uma estrutura semiósica ordenadora e invisível aos usuários. Isso ocorre, em parte, devido à instantaneidade e prontidão das respostas da IA, bem como à naturalização da sua razoabilidade, que tendem a minimizar a necessidade de uma crítica mais aprofundada.

O estudo visa contribuir para o debate sobre racismo algorítmico, expondo as implicações éticas das semioses algorítmicas. Tal enfoque possibilita uma análise crítica de como os sistemas de IA generativa de imagens não apenas refletem, mas também moldam as construções sociais de identidade, alteridade e raça. Além disso, contribui para o debate sobre ética e IA, bem como salienta a necessidade de abordar o racismo em sua dimensão algorítmica, a fim de encontrar soluções para mitigar o enviesamento racial e seus efeitos ao evidenciar a falsa neutralidade das semioses maquínicas.

Este artigo está estruturado da seguinte maneira: a primeira seção, Introdução, aborda os elementos contextuais que originaram o estudo, bem como os seus pressupostos, questão norteadora, objetivos e justificativa. A segunda seção, Referencial teórico, contempla toda a base teórica utilizada. A terceira seção, Metodologia, caracteriza a pesquisa conforme a sua natureza, abordagens empregadas, recorte empírico, instrumentos e técnicas de coleta e análise de dados. A quarta seção, Resultados, apresenta as principais descobertas e reflexões a partir da análise dos dados obtidos. Por

fim, a quinta seção, Considerações Finais, dialoga com os objetivos inicialmente propostos, resume os resultados obtidos e apresenta reflexões sobre as implicações da pesquisa, bem como sobre os possíveis estudos futuros.

2 REFERENCIAL TEÓRICO

2.1 Inteligência Artificial Generativa: breve histórico, recortes e definições

A IA generativa é uma área da Inteligência Artificial voltada para a criação de conteúdos que têm a intenção de ser inovadores e originais. Para alcançar esse objetivo, os sistemas analisam as características dos conteúdos para os quais foram projetados, utilizando uma grande quantidade de exemplos reais, geralmente sem a necessidade de supervisão, e conseguem gerar novos conteúdos com essas mesmas propriedades (Corredera, 2023).

Ferramentas baseadas em IA generativa processam grandes volumes de dados a fim de identificar padrões e gerar novas composições, utilizando técnicas algorítmicas avançadas. Embora os mais recentes aprimoramentos dessas tecnologias tenham se destacado, suas raízes remontam aos anos 1940, com a proposta do neurônio artificial de McCulloch-Pitts. No entanto, foi a introdução de modelos como as Redes Generativas Adversariais (GANs), em 2014, e o *Transformer*, em 2017, que impulsionou a IA generativa para novos patamares de sofisticação, culminando no desenvolvimento de modelos como o GPT (*Generative Pre-trained Transformer*), capaz de criar conteúdos como textos e imagens com notável realismo.

Destaca-se em Corredera (2023) um ponto central da IA generativa: sua capacidade de criação de novos conteúdos com base em dados preexistentes. Essa abordagem sublinha a natureza desses sistemas, que conseguem identificar padrões e características oriundos de exemplos prévios para gerar novos resultados. No entanto, a noção de "novo e original" pode ser debatida, pois, apesar de parecerem inovadores, esses conteúdos são baseados em dados anteriores, o que tem suscitado questões sobre criatividade genuína e o limite entre reprodução e inovação. Além disso, o uso de exemplos "reais" reflete uma dependência de dados que, se enviesados ou limitados, podem impactar a qualidade e diversidade dos conteúdos gerados. Isso reforça a importância de examinar cuidadosamente os conjuntos de dados usados para garantir que a produção seja

verdadeiramente representativa e não simplesmente a amplificação e replicação de padrões.

A IA generativa utiliza um modelo computacional conhecido como redes neurais artificiais. Conforme Haykin (2001) e Baeza-Yates e Ribeiro-Neto (2011), uma rede neural artificial é uma estrutura computacional que se inspira na organização e no funcionamento do cérebro humano, sendo desenvolvida para identificar padrões em dados por meio de um algoritmo de aprendizagem. A definição desses autores ressalta a analogia entre esses sistemas e o modelo de funcionamento do cérebro humano, destacando o aspecto de inspiração biológica que permeia a arquitetura dessas redes. Ao enfatizarem que as redes neurais são projetadas para reconhecer regularidades a partir de dados, os autores destacam a função central dessas redes no “aprendizado” da máquina: identificar correlações e características relevantes em grandes volumes de dados de forma automatizada. O uso de um algoritmo de aprendizagem, mencionado na definição, é fundamental para que as redes possam se aprimorar ao longo do tempo, refinando sua capacidade de classificação e previsão.

O modelo de McCulloch e Pitts (1943) é considerado um marco fundamental no desenvolvimento das redes neurais artificiais e, por extensão, da própria Inteligência Artificial. Esse modelo, embora rudimentar comparado às redes neurais modernas, foi revolucionário ao demonstrar que operações lógicas e funções matemáticas poderiam ser realizadas por um conjunto de neurônios artificiais interconectados, antecipando a lógica de aprendizado e processamento que veio a ser desenvolvida em modelos posteriores. O trabalho de McCulloch e Pitts (1943) lançou as bases para o que hoje se denomina redes neurais profundas, que são estruturas complexas com várias camadas de neurônios artificiais capazes de processar dados de forma muito mais sofisticada. Apesar das limitações do modelo, como a ausência de mecanismos de aprendizado, ele forneceu o primeiro passo na exploração de como redes de neurônios artificiais poderiam emular, em certa medida, funções cognitivas do cérebro humano.

Durante as décadas de 80 e 90, as redes neurais artificiais começaram a atrair mais atenção, especialmente com a criação do algoritmo de retropropagação, por Geoffrey Hinton, David Rumelhart e Ronald J. Williams, em 1986. Esse algoritmo é um método essencial para o treinamento das redes neurais artificiais, pois ajusta os pesos das conexões entre os neurônios artificiais, com o objetivo de reduzir o erro entre as previsões da rede e os valores reais conhecidos (Rumelhart; Hinton; Williams, 1986), (Hinton, 1992).

No entanto, naquela época, os estudos enfrentavam limitações relacionadas aos dados disponíveis e à capacidade computacional insuficiente.

A partir dos anos 2000, com o avanço da capacidade de processamento dos computadores e a maior disponibilidade de volumes massivos de dados, Bengio, LeCun e Hinton (2021) ressaltam que as pesquisas sobre redes neurais artificiais e aprendizagem de máquina (*machine learning*) ganharam novo impulso, agora sob o conceito de redes neurais profundas (*deep learning*). Esses autores afirmam que o desempenho dos sistemas de IA generativa melhora conforme aumentam, tanto a quantidade de dados disponíveis, quanto a robustez dos recursos computacionais. Assim, quanto maior for o volume e a diversidade dos dados - e maior a capacidade de processamento - mais eficaz tende a ser o treinamento dos sistemas de IA.

Goodfellow *et al*. (2014) apresentaram as Redes Generativas Adversariais (GANs), que introduziram uma técnica mais aprimorada para o treinamento desses sistemas, levando à geração de conteúdos extremamente realistas. Essa inovação resultou em avanços significativos na produção de imagens, vídeos e outros tipos de conteúdo.

Destaca-se ainda o desenvolvimento do modelo conhecido como *Transformer*, criado por Vaswani *et al.* (2017). Essa contribuição científica introduziu uma nova arquitetura de rede neural que possibilitou a criação de modelos de linguagem, capazes de gerar, adaptar e manipular texto em linguagem natural, como o GPT (*Generative Pre-trained Transformer*) da OpenAI. Os *Transformers* tornaram-se fundamentais para a geração de texto e a tradução automática.

Nos anos seguintes, a empresa OpenAI lançou várias versões dessa tecnologia: o GPT-1 em 2018, seguido pelo GPT-2 em 2019, o GPT-3 em 2020 e o GPT-4 em 2023. De maneira geral, esses modelos apresentaram uma considerável capacidade de gerar conteúdos informacionais contextualmente relevantes. Além disso, outras plataformas e serviços de IA generativa também se destacaram nesse contexto.

Um exemplo importante para esta pesquisa é o DALL-E, que é uma extensão da série GPT e emprega técnicas de *deep learning* para interpretar comandos (*prompts*) a partir de textos e criar representações visuais, o que propicia a geração de imagens com uma variedade de estilos artísticos e conteúdos, que vão desde cenas e objetos cotidianos até cenas e objetos complexos e imaginários.

Os sistemas de IA generativa de imagem atuais se baseiam na combinação de técnicas mais específicas de PLN e *machine learning*, como o CLIP (*Contrastive Language*

Image Pre-training) e em avanços no contexto da *machine learning,* como os modelos de difusão (*Diffusion Models*). O CLIP, na concepção de Radford *et al* (2021) é um modelo que estabelece associação entre o signo textual e o signo imagético, sendo treinado mediante pares de texto-imagem extraídos da web. Desse modo, não apenas associa palavras e imagens, mas também reproduz as representações culturais presentes em seus dados de treinamento.

Já os modelos de difusão operam de forma a gerar imagens a partir de um processo iterativo. Ho, Jain e Abbeel (2020) introduziram uma técnica que adiciona ruído (alterações aleatórias nos *pixels* que distorcem o conteúdo visual) a uma imagem de maneira progressiva durante o treinamento (processo de difusão) e ensina o modelo a reverter esse ruído a fim de refinar a imagem para aproximá-la de uma saída que melhor se alinha ao *prompt* fornecido. Esse processo busca aprimoramentos, tanto na qualidade visual, quanto na preservação da coerência semântica com o que é solicitado no *prompt*. A abordagem de Ho, Jain e Abbeel (2020) inspirou uma nova onda de pesquisas em modelos de *machine learning* baseados em difusão, consolidando-os como padrão atual para tarefas de geração visual com níveis consideráveis de qualidade, realismo e precisão.

Conforme apontam Luccioni *et al.* (2023) sistemas de IA generativa de texto e de imagens codificam preconceitos de formas diferentes e a literatura atual ainda não possibilita afirmar se ocorre amplificação ou agravamento do enviesamento de modo mais intenso em uma ou outra modalidade.

Entretanto, Moura e Braga (2023) investigaram a representatividade feminina em ferramentas que empregam IA generativa para a criação, tanto de textos, quanto de imagens e observaram em seus resultados elementos como a erotização e a sensualização do corpo, com destaque para o corpo das mulheres negras, nas ferramentas dedicadas a produção de imagens. Apontam que resultados similares não foram encontrados nas ferramentas dedicadas a geração de conteúdo textual. Os referidos autores evidenciaram especificamente as disparidades na forma como mulheres negras e brancas são representadas, ressaltando que essas diferenças não decorrem de características físicas explícitas nos *prompts* de criação, mas sim dos vieses enraizados nos dados de treinamento.

Embora autores como Castro, Otero e Rolán (2023); Peñaranda (2023); De Zárate *et al.* (2024) abordem questões relevantes sobre preconceitos e vieses nos Modelos de Linguagem de Grande Escala (LLM), cujo principal expoente é o ChatGPT, o foco do

presente trabalho é o enviesamento nos sistemas de IA generativa de imagens, abarcando então o contexto que compreende os modelos denominados multimodais (MLLM).

De acordo com Ibañez (2024) as pesquisas que abordam o enviesamento em modelos e ferramentas de IA generativa, focados na criação de imagens a partir de texto, são recentes e ainda não volumosas. Revelam, em sua maioria, vieses consideráveis nas dimensões ocupacionais, de raça e de gênero.

2.2 O conceito de imagem à luz da caixa-preta flusseriana

No campo da Ciência da Informação a imagem é abordada como uma representação visual que pode ser descrita, analisada e organizada a partir de várias perspectivas e dimensões. Trata-se de um conteúdo informacional complexo, devido ao seu amplo potencial semiósico e sendo assim, capaz de desencadear diferentes níveis de interpretação e significação, que variam de acordo com quem a observa, do contexto em que é analisada e dos elementos que a constituem, conforme evidenciam Smit (1986); Shatford (1994); Manini (2001). Isso ocorre porque além dos aspectos concretos e literais, a imagem contém camadas abstratas e simbólicas que podem ser interpretadas de diversas maneiras, de acordo com os contextos históricos, sociais e culturais em que está inserida.

Mas o que diferencia as imagens tradicionais (criadas diretamente pelo ser humano, a partir de sua percepção do mundo) das imagens geradas algoritmicamente? As reflexões que confluem para possíveis respostas a esse questionamento encontraram terreno fértil no pensamento do filósofo Vilém Flusser (1920-1991).

Para Flusser (1985) a imagem é uma mediação entre o ser humano e o mundo, que propicia um espaço interpretativo a quem a contempla, dado o seu caráter conotativo. Na concepção flusseriana o ato de codificar e decodificar o real em superfícies bidimensionais é central para a definição do conceito de imagem, mas envolve a mediação de um sujeito capaz de atribuir significado aos símbolos criados.

As IA generativas, embora capazes de criar imagens convincentes e inovadoras, não realizam esse processo da mesma forma que os seres humanos, pois seguem um conjunto de instruções e probabilidades baseadas em padrões previamente modelados nos dados de treinamento.

O que as IA generativas fazem poderia ser compreendido, portanto, como uma "simulação" da imaginação humana. Elas processam grandes volumes de dados, identificam padrões e geram combinações criativas, mas fazem isso de maneira mecânica,

o que implica que a intencionalidade e a subjetividade que caracterizam a imaginação humana não devem influenciar esse processo, nem mesmo de modo indireto, visto que, não há, na IA generativa, a capacidade de experimentar e interpretar o mundo em quatro dimensões para depois abstraí-lo em símbolos, pois operam a partir de cálculos estatísticos, e suas criações resultam de uma síntese algorítmica, não de uma experiência direta do real.

Em seu conceito de imagem técnica, Flusser (1985, p.10) nos traz a definição de que:

> [...] trata-se da imagem produzida por aparelhos. Aparelhos são produtos da técnica que, por sua vez, é texto científico aplicado. Imagens técnicas são, portanto, produtos indiretos de textos, o que lhes confere posição histórica e ontológica diferente das imagens tradicionais (Flusser, 1985, p.10).

Nesse contexto, as imagens técnicas são aquelas que resultam de aparatos mecânicos, como as câmeras fotográficas, que mediam a relação entre o ser humano e o mundo. Com a IA generativa nos deparamos com uma nova forma de imagem técnica, com um grau ainda maior de abstração e mediação. O aparelho, agora, não apenas captura o mundo visual de maneira programada, mas também é capaz de gerar imagens a partir de algoritmos e volumes exponenciais de dados, sem uma representação direta do real. Essas novas imagens técnicas se desconectam ainda mais do referente original. Se antes a imagem técnica já era uma abstração, com a IA generativa, essa abstração se torna uma simulação de algo que nunca existiu no mundo tangível.

A partir de Flusser (1985) entende-se que a codificação da realidade por meio de algoritmos pode ser vista como um novo nível de programação do imaginário, onde as imagens criadas pela IA não representam apenas o mundo físico, mas constroem novos mundos possíveis, repletos de signos e significados que transcendem a experiência humana direta.

Argumenta-se ainda que as IA generativas apresentam ao ser humano uma nova camada de mediação e construção da realidade, visto que “ontologicamente as imagens tradicionais imaginam o mundo; as imagens técnicas imaginam textos que concebem imagens que imaginam o mundo” (Flusser, 1985, p.10). Essas tecnologias estabelecem então um novo paradigma de mediação, em que a geração de imagens e outros conteúdos não se limita mais à reprodução ou representação do mundo físico, mas à criação de realidades que mesclam dados, interpretações e simulações baseadas em padrões.

Assim, o papel da IA vai além de simplesmente mediar a realidade pois possibilita reconfigurar e reinterpretar o que se entende como real. As imagens geradas por IA não são apenas representações de algo que já existe, mas podem compor novas realidades visuais ou simbólicas, criadas a partir de processos algorítmicos complexos.

Por fim, argumentamos que a ideia de Flusser (1985) sobre a necessidade de se esclarecer a caixa preta da qual se originam as imagens técnicas se aplica de forma relevante às IA generativas, mesmo que o filósofo não estivesse se referindo a elas diretamente. A ideia central é que, quando observamos a lógica de "*input*" e o "*output*" das máquinas, estamos apenas enxergando os resultados finais — ou seja, as imagens — e não compreendendo o que realmente acontece dentro dos processos técnicos que geram essas imagens. Esse processo oculto é a "caixa preta" que ele menciona, uma metáfora para os mecanismos internos das tecnologias, que permanecem invisíveis e incompreensíveis para a maioria das pessoas.

No contexto das IA generativas, a "caixa preta" se torna ainda mais complexa. Os algoritmos de *machine learning* e as redes neurais que geram imagens (e outros conteúdos) funcionam de maneiras que muitas vezes são difíceis de decifrar, mesmo para especialistas. O processo de treinamento de uma IA envolve ajustamento de parâmetros e inúmeras camadas de processamento que não são diretamente acessíveis ou compreensíveis pelo usuário comum.

Portanto, assim como Flusser (1985) aponta para o "analfabetismo" em relação às imagens técnicas, podemos dizer que a sociedade contemporânea enfrenta um tipo de analfabetismo em relação aos conteúdos informacionais gerados por IA. Apenas ver o produto final, sem compreender o funcionamento interno dessas tecnologias deixa a sociedade em uma posição passiva diante de tais conteúdos, sem a capacidade crítica de decifrar sua origem, seu significado e os vieses envolvidos no seu processo de geração.

2.3 Semiose e IA generativa

A semiótica de base peirciana constitui um campo teórico capaz de auxiliar na compreensão dos processos de produção de sentido mediados por dispositivos tecnológicos, tanto na assistência quanto na tradução dos processos semiósicos maquínicos.

Nesse contexto, é importante destacar que, na atualidade, ainda não é possível considerar a autonomização total desses processos, especialmente porque eles dependem

da intervenção humana em diversas etapas, baseando-se em uma concepção de semiose de caráter orientador.

Segundo Moura (2002, p. 52), a concepção semiósica refere-se à consciência que o homem tem da ação de seu interpretante na mente de outro e ao movimento, consciente ou inconsciente, que ele faz para influenciar essa outra mente. Trata-se da maneira como um sujeito tenta intervir na semiose de outrem por meio da interação do intérprete com uma estrutura semiósica construída por ele. Nesse processo, o interpretante é moldado também pela concepção semiósica que orienta tal intervenção.

Ao teorizar sobre essas intervenções nos processos perceptivos alheios, Charles Peirce, segundo a autora, não se preocupava em explorar a dimensão moral desse fenômeno. Seu foco estava, sobretudo, em posicionar o homem como signo, intérprete e gerador de novos signos, enfatizando que toda percepção é, em essência, uma forma de interpretação.

A tão proclamada abertura das obras contemporâneas à intervenção criativa do receptor ainda enfrenta limitações impostas pelo projeto semiósico concebido em sua gênese.

Para Peirce (CP 5484) a semiose, decorre da influência trirrelativa que envolve o signo, o objeto e o interpretante. Tal perspectiva, nos auxilia na compreensão do caráter fugidio do signo que projeta sempre para o futuro, a consecução do sentido.

Segundo Pinto (1995), o signo:

> […] cria significação, em vez de passivamente esperar que o sujeito o invista de sentido. Em outras palavras, o sujeito interpreta o signo à sua maneira e gera nesse processo seus próprios interpretantes, mas o signo não é vazio, e o sujeito não o preenche através de um fiat divino (Pinto, 1995, p. 50).

O interpretante, por sua vez, é entendido como uma entidade objetiva do signo, resultante da semiose. Ele é responsável pelo processo de significação e por seu direcionamento dinâmico rumo a um futuro final que nunca se encerra.

A perspectiva trirrelativa do signo pode oferecer uma compreensão mais aprofundada do processo semiósico que ocorre no contexto da IA generativa. Nela, a semiose algorítmica se concretiza por meio dos *prompts*, que conectam o projeto semiósico, que orienta as funcionalidades do dispositivo tecnológico, à etapa responsiva gerenciada pelos usuários na cadeia semiósica. Embora as respostas sejam frequentemente percebidas como instantâneas, o interpretante algorítmico gerado é dependente de vários fatores, como o banco de dados utilizado no treinamento, as visões

de mundo e os objetivos incorporados ao projeto, além da experiência colateral dos usuários na formulação de *prompts* e no reconhecimento ou estranhamento da razoabilidade das respostas obtidas.

É importante ressaltar que o funcionamento detalhado dos algoritmos que possibilitam essa semiose é cada vez mais considerado um segredo industrial, escapando deliberadamente à observância e análise de seus usuários finais.

Os algoritmos, que compõem parte de nossa preocupação, integram os ecossistemas informacionais e funcionam como repertórios de procedimentos codificados que orientam a consecução de uma tarefa simples ou complexa. Envolvem, segundo Gillespie (2018) a proposição de padrões de inclusão, ciclos de antecipação, avaliação de relevância, a promessa de objetividade algorítmica, entrelaçamento com a prática e a produção de públicos calculados.

Os padrões de inclusão abrangem as definições operacionais e os propósitos do algoritmo, assim como suas fontes de informação e conexão. É nessa etapa que se estabelecem efetivamente os critérios de inclusão e exclusão. Os ciclos de antecipação envolvem o processamento e a indexação ativa das informações dos usuários, permitindo que o sistema, em que o algoritmo opera, forneça sempre respostas pertinentes, específicas e em tempo real às futuras demandas.

A avaliação de relevância e a promessa de objetividade se relacionam à análise dos propósitos para os quais o algoritmo foi criado. Segundo Gillespie (2018), os critérios adotados nesse processo são uma miríade de parâmetros que determinam o que será apresentado como resposta a uma demanda dos usuários.

Mais do que meras ferramentas, os algoritmos também são estabilizadores da confiança, garantias práticas e simbólicas de que suas avaliações são justas e precisas, livres de subjetividade, erro ou tentativas de influência:

> [...] acima de qualquer coisa, os provedores dos algoritmos de informações devem assegurar que seus algoritmos são imparciais. A performance da objetividade algorítmica tornou-se fundamental para a manutenção dessas ferramentas como mediadoras legítimas do conhecimento relevante (Gillespie, 2018, p. 107-108).

A produção de públicos calculados permite que os propositores das aplicações mantenham e organizem a conectividade, cada vez mais restrita, entre os públicos e seus interesses expressos no ecossistema informacional e buscam, nesse sentido, identificar, sistematizar e modelar os padrões de comportamento informacional de tais públicos.

2.4 Maquinarias, dispositivos e vieses no contexto da adoção da IA generativa

O crescente uso de tecnologias baseadas em IA generativa, em situações cotidianas, científicas, ou no entretenimento, viabilizou uma miríade de novos produtos e serviços que atuam de forma sistemática no encurtamento dos processos de produção de sentido. Isso ocorre porque seu caráter assistivo traz, como tendência, a dinamização do processo interpretativo a partir da oferta de suprimentos regulares de sentido. Em tais circunstâncias, observa-se uma forte tendência ao exercício do poder, à medida que os interesses em jogo são subsumidos por uma camada de alta tecnicidade.

Nesse contexto, torna-se pertinente problematizar a incorporação da branquitude como modelo hegemônico de humanidade, replicado pelas IA generativas de imagens. Essas tecnologias tendem a viabilizar um projeto político de assujeitamento que promove a constante morte social dos sujeitos considerados não hegemônicos, produzindo, assim, uma 'alucinação consensual' cujos desdobramentos incluem a hierarquização racial.

Isso se deve, em grande parte, ao legado escravocrata de nossa sociedade, historicamente organizada com base em uma economia consolidada pela comodificação de corpos racializados como estratégia de dominação, subjugação e enriquecimento. Nesse cenário, a branquitude é compreendida como uma tecnologia de poder, consolidada no processo de colonização, que se manifesta em "práticas discursivas, formas de conhecimento e linguagem". Essas práticas tendem a se mobilizar como um pacto e uma estratégia de controle social, orientando sentidos, acessos e naturalizando violências.

O pacto narcísico da branquitude, Conforme Bento (2022), é uma aliança tácita silenciosa, que se consolidou historicamente com o propósito de naturalizar a hierarquia racial do privilégio e do mérito e salvaguardar, por transmissão geracional negativa e construção simbólica, o poder de definir quem pertence ao grupo 'de dentro' e quem está 'de fora'.

O circuito de apequenamento dos sujeitos racializados é aprofundado pelo que Collins (2019) denominou de 'imagens de controle'. Essas imagens operam pela inculcação de autoimagens moldadas por matrizes de dominação. Nesses casos, a consolidação de uma autoimagem medíocre, imposta por fontes e agentes externos ao sujeito imaginado, é naturalizada.

Carneiro (2023) nomeia essas maquinarias, voltadas para a hierarquização e invisibilização, como dispositivos de racialidade. Para a autora, o caráter de dispositivo

permite a sua operacionalização extemporânea, destacando o uso estratégico e o efeito ontológico que estabelece o caráter enunciativo, em que a brancura se configura como norma, antagonizando-se com o 'outro', demarcado por uma ontologia da diferença.

No campo da admissibilidade do conceito de branquitude e seus efeitos sociais, nota-se uma forte conexão entre os dispositivos de racialidade e seus desdobramentos, que se manifestam por meio das imagens de controle. Esses mecanismos operam também nos ambientes digitais, facilitando a ocorrência do racismo algorítmico, conforme abordado por Silva (2019; 2022). Nesse cenário, a conectividade da vida cotidiana e a racialização das relações permitem o surgimento de vieses resultantes de um olhar codificado e manipulado por algoritmos.

No caso específico da produção, interpretação e circulação de imagens, a área de computação visual exerce um papel central ao promover a gestão do visível, baseada no monitoramento e na criação de padrões. Esse processo pode tanto gerar a superexposição de sujeitos negros quanto contribuir para o apagamento desses corpos, revelando a dualidade do tratamento algorítmico em relação à visibilidade e invisibilidade racial.

O atravessamento promovido pelos algoritmos influencia diretamente a conformação da percepção dos sujeitos em relação às representações culturais, tanto por meio das respostas imagéticas quanto pela produção de um imaginário artificial. Esse imaginário, por sua vez, tende a reproduzir as lógicas de naturalização das hierarquias de opressão, oferecendo como resposta, imagens — sejam elas existentes ou sintéticas — que reforçam essas hierarquias, contribuindo para a sua perpetuação.

3 METODOLOGIA

Este estudo se caracteriza como exploratório e quali quantitativo. Aborda uma temática sobre a qual ainda se observa baixa produção científica crítica no Brasil e considera tanto a importância de uma abordagem interpretativa dos dados, quanto a sua quantificação a fim de que sejam analisados de forma a identificar padrões, temas e narrativas predominantes. Trata-se ainda de uma perspectiva metodológica que combina a análise semiótica e a análise crítica de conteúdo.

Os experimentos que originaram os dados da pesquisa foram compostos da seguinte maneira:

Etapa 1 - Identificação e seleção de ferramentas de IA generativa, parcialmente ou totalmente gratuitas, que realizem a criação de imagens a partir de *prompts* (comandos)

textuais. Nessa etapa foi utilizado o diretório *Insidr.AI*[1], que mapeia e categoriza sistemas e aplicações de IA disponíveis na web. Foram selecionadas 10 ferramentas.

Etapa 2 - Elaboração dos *prompts* para a geração das imagens a partir de critérios temáticos e semióticos. O texto de descrição aplicou o primeiro nível do modelo de Shatford (1986), composto pelas categorias "DE genérico" e "DE específico". Foram elaborados e testados oito *prompts* que são atinentes às seguintes temáticas: a) identidade feminina; b) identidade masculina; c) identidade feminina negra; d) identidade masculina negra e e) dimensão laboral.

Etapa 3 - Aplicação dos *prompts*, de forma aleatória, no universo composto pelas 10 ferramentas selecionadas. Elaboração, organização e armazenamento de um acervo imagético digital composto por um total de 155 imagens, geradas a partir desses *prompts*.

Etapa 4 - Categorização e análise quantitativa e qualitativa das 155 imagens a fim de identificar vieses e padrões. Descrição e análise de um recorte constituído por 47 imagens.

Os oito *prompts* aplicados neste estudo, foram formulados em inglês e foram caracterizados por um grau crescente de especificidade, conforme o esquema abaixo:

1) "*a man*"
2) *"a woman"*
3) *"a black man"*
4) *"a black woman"*
5) *"a black man working"*
6) *"a black woman working"*
7) *"a brazilian black man working"*
8) *"a brazilian black woman working"*

Tal abordagem empregou uma lógica reversa do modelo de *Sara Shatford*, para criar *prompts* e não para analisar as imagens geradas pela IA. Ao aplicar apenas as categorias objetivas: "DE genérico" e "DE específico", na formulação dos *prompts*, buscou-se intencionalmente controlar o nível de detalhamento semântico fornecido.

O Quadro 1 apresenta todas as ferramentas de IA generativa utilizadas na pesquisa.

Quadro 1: caracterização das ferramentas de IA generativa.

	Ferramenta	Ano de criação	Empresa	País	Modelo	Acesso
1	BlueWillow	2022	LMWR Technology	EUA	bluewillow	Parcialmente gratuito
2	Canva	2012	Canva Pty Ltd	Austrália	DALL-E	Parcialmente gratuito
3	Bing Image Creator	2023	Microsoft	EUA	DALL-E3	Parcialmente gratuito
4	Crayon	2022	Craiyon LLC	EUA	DALL·E mini	Gratuito
5	FreePik	2010	Freepik Company	Espanha	NI	Parcialmente gratuito
6	Getimg.ai	2022	Webrockets Sp	Polônia	getimg.ai	Parcialmente gratuito
7	ImagineArt	2023	Vyro LLC	EUA	NI	Parcialmente gratuito
8	Leonardo AI	2022	Leonardo Interactive Pty Ltd	Austrália	Leonardoai	Parcialmente gratuito
9	Magic Studio	2020	Aarzoo, Inc	EUA	NI	Parcialmente gratuito
10	Runway	2018	Runway AI, Inc	EUA	NI	Parcialmente gratuito

Fonte: Dados da pesquisa

No que concerne aos critérios adotados para a seleção das 10 ferramentas empregadas neste estudo: a) todas possuem a funcionalidade de criação de imagens a partir de *prompts* textuais; b) possibilitam que os usuários as utilizem de forma gratuita, ou parcialmente gratuita, a partir de versões básicas ou com um acesso limitado a um determinado número de imagens possíveis de serem geradas; c) permitem a criação de imagens caracterizadas por diferentes estilos artísticos, desde ilustrações até fotos verossímeis, bem como ajustes de parâmetros como o nível de detalhe, cor, iluminação e composição; d) estimulam o *feedback* dos usuários sobre as imagens geradas a fim de melhorar a qualidade e a relevância das criações; e) aplicam modelos próprios para a geração das imagens ou empregam modelos proprietários, como o DALL-E, embora algumas não sejam suficientemente transparentes quanto ao modelo aplicado, nesses casos foram sinalizadas como o indicador NI (Não Identificado) no Quadro 1.

Algumas ferramentas estão associadas a plataformas mais abrangentes, que disponibilizam uma gama de serviços, produtos e instrumentos para a criação de outros tipos de conteúdos, não se restringindo à imagens. Nesse sentido, *FreePik* e *Canva* se destacam por serem plataformas mais amplas, oferecem produtos, serviços e recursos gráficos, que incluem ilustrações, imagens e modelos. Contudo, uma das funcionalidades mais recentes disponíveis nessas plataformas possibilita a geração de imagens por IA a partir de *prompts* textuais, o que atendeu ao principal critério aplicado pelo estudo.

Na próxima seção, serão analisadas situações de interação e uso assistivo de ferramentas de produção de imagens com IA, com o objetivo de compreender seu

funcionamento, suas limitações e como os vieses são operacionalizados pela IA generativa. Como não é possível inserir e analisar todas as imagens que compõem o *corpus* da pesquisa, no presente artigo, optou-se por elaborar apenas alguns quadros com justaposição das imagens para fins de demonstração e exemplificação.

4 RESULTADOS

O método desenvolvido possibilitou a compreensão da tecnicidade na produção de imagens de síntese, seus desdobramentos em termos de operacionalidade algorítmica, e a produção de sentidos. Esse processo foi analisado no contexto do uso assistivo de IAs para a criação de imagens, considerando a semiose prevista nas interações mediadas pelos *prompts*. Embora os *prompts* tenham sido organizados com expressões comuns e convencionais, os resultados evidenciaram uma tendência à reprodução de vieses classistas e racistas, tal como apontado por Ibañez (2024). Esses vieses funcionam como continuidades de valores arraigados, que orientam a semiose inscrita em pactos hierárquicos, os quais, em vez de se autocorrigirem por meio das interações, acabam se sobrepondo às realidades objetivas.

A semiose algorítmica refere-se ao processo de criação de significados em que os signos envolvem a interação entre o algoritmo, os dados de entrada (*prompts*), os dados de treinamento e o contexto cultural em que esses elementos estão inseridos. Esse tipo de semiose viabiliza tecnicamente um projeto de significação e, por consequência, o enviesamento algorítmico. A primeira abarca como os significados são gerados e interpretados, enquanto o segundo aborda as distorções que podem afetar esse processo.

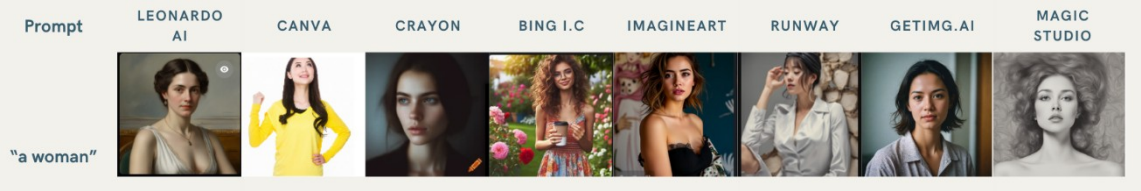
Observou-se que, a partir da aplicação de *prompts* mais genéricos, as ferramentas majoritariamente geram imagens de pessoas brancas de classe média alta (90,9% para homens e 92% para mulheres). Isso revela a naturalização de um padrão visual e estético para os seres humanos.

Das 11 imagens geradas com o *prompt* “*a man*”, em quatro ferramentas diferentes, 10 retrataram homens brancos de classe média alta. Das 26 imagens geradas a partir do *prompt* “*a woman*” em oito diferentes ferramentas, 24 imagens retrataram mulheres brancas de classe média alta.

Tal resultado, assinala uma perspectiva alinhada ao funcionamento do dispositivo de racialidade que tende a naturalizar a branquitude e a classe social média ou alta como a

norma nos resultados obtidos. O Quadro 2 apresenta um recorte de dados com imagens geradas a partir do *prompt* “*a woman*”.

Quadro 2: extrato de imagens geradas com o *prompt* “*a woman*”.

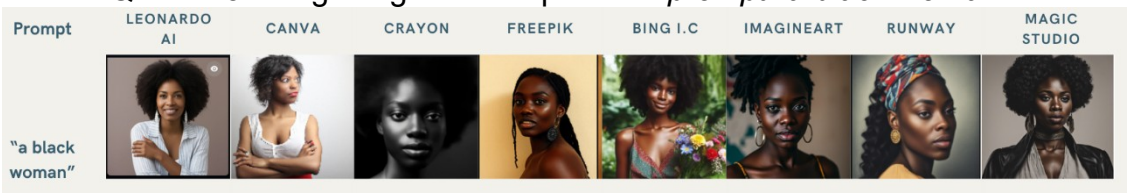


Fonte: Dados da pesquisa.

A ausência de diversidade racial observada nos resultados reflete a hegemonia de um único grupo étnico-racial na produção das imagens, evidenciando padrões normativos pautados pela branquitude. Esses padrões indicam que, quando não há especificação racial, as IA generativas analisadas majoritariamente tendem a reproduzir o que é socialmente instituído como “normal” e “universal”, reforçando a normatividade supremacista de um determinado grupo em detrimento de outro.

A partir da aplicação do *prompt* “*a black man*”, em seis ferramentas diferentes, foram geradas 16 imagens. Das quais, 11 retraram homens negros de classe média baixa. 100% dos indivíduos eram homens negros de pele preta retinta e cabelos crespos. Quando aplicado o *prompt* “*a black woman*” em nove ferramentas diferentes, das 29 imagens geradas, 22 eram de mulheres negras de pele preta retinta. Sendo que 100 % delas foram retratadas com os cabelos crespos ou encaracolados, conforme demonstra o Quadro 3.

Quadro 3: imagens geradas a partir do *prompt* “*a black woman*”.



Fonte: Dados da pesquisa.

Percebeu-se que quando o *prompt* especifica *“black man”* e “*black woman*”, há uma mudança nas características socioeconômicas das imagens geradas. Homens negros e mulheres negras são frequentemente associados a imagens de indivíduos de baixa renda (68,7% dos homens e 75% das mulheres negras), reforçando os estereótipos raciais ligados à classe. No caso das mulheres negras, a maioria das imagens as representa com cabelos crespos ou encaracolados, reforçando um imaginário padronizado de "autenticidade" da negritude. Embora o cabelo seja um marcador importante de identidade cultural, a ausência

de diversidade nas representações indica o reducionismo da identidade negra a um conjunto limitado de características visuais.

Do total de 24 imagens geradas a partir do *prompt* "*a black man working*", aplicado a oito ferramentas diferentes, 18 (75%) remeteram a algum tipo de atividade laboral braçal ou manual e 12 (50%) remeteram a homens negros realizando algum tipo de trabalho estritamente braçal. Sendo que 14 imagens (58%) remetem a homens negros de baixa renda. Em nove imagens (37,5%) o indivíduo aparece trajando um avental. O Quadro 4 exemplifica a caracterização dos dados observados:

Quadro 4: imagens geradas a partir do *prompt* "*a black man working*"

Fonte: Dados da pesquisa.

Quando utilizado um *prompt* ainda mais específico: "*a brazilian black man working*", das 12 imagens geradas em três ferramentas diferentes, nove imagens (75%) remetem a atividades estritamente braçais e 100% retrataram homens negros de baixa renda em atividades braçais ou manuais.

Já do total de 26 imagens geradas com o *prompt* "*a black woman working*" em oito ferramentas diferentes, nove imagens (34%) retrataram mulheres negras de baixa renda e oito imagens (30,7%) retrataram a mulher negra também trajando um avental. Quatro (15,3%) dessas imagens retrataram trabalhos estritamente braçais. Ao utilizar o *prompt* mais específico: "*a brazilian black woman working*", das 11 imagens geradas por três ferramentas diferentes, nove (81,8%) eram imagens de mulheres negras de baixa renda e quatro (36%) retratavam um trabalho estritamente braçal, duas dessas quatro imagens retrataram mulheres negras trajadas com um avental. O Quadro 5 constitui um extrato que possibilita exemplificar as características presentes nos dados analisados.

Quadro 5: imagens geradas a partir do *prompt* "*a brazilian black woman working*".

Fonte: Dados da pesquisa.

Destaca-se a recorrência de imagens de trabalho manual ou braçal associadas às pessoas negras, principalmente a homens negros. Esses dados sugerem que a IA replica estereótipos visuais que relacionam a negritude a ocupações de baixo *status*, o que reforça a atuação dessas ferramentas no âmbito do que Carneiro (2023) denominou dispositivo de racialidade, compondo uma miríade de discursos e práticas que configuram o racismo e a discriminação racial e entre outros efeitos, impõem limites à mobilidade social de pessoas negras, por meio de representações simbólicas. A associação de pessoas negras com trabalhos manuais ou braçais, especialmente o uso de avental, reproduz estereótipos históricos de servidão e subalternidade.

As correlações automáticas entre negritude, baixa renda e trabalho manual ou braçal, revelam que os *datasets* utilizados para treinar essas ferramentas estão profundamente marcados por estereótipos raciais e de classe. Os resultados demonstram que as ferramentas inferem, replicam e até amplificam as desigualdades estruturais presentes na sociedade e apenas em exceções ofereceram representações diversificadas ou igualitárias.

Flusser (1985) alertava para o risco de desumanização nas imagens técnicas, já que elas impõem uma camada tecnológica entre a experiência direta e a representação visual. No caso das IA generativas, há um distanciamento ainda maior entre o humano (pessoas negras reais em toda a sua diversidade) e as imagens geradas por algoritmos. A desumanização se manifesta não apenas no conteúdo estereotipado das imagens, mas também no próprio processo pelo qual essas imagens são criadas. Sob esse aspecto, as IA generativas analisadas contribuem para uma visão desumanizadora da negritude, que reduz pessoas pretas a padrões simplificados - e simplificadores - e distorce sua complexidade histórica, cultural e social. Enquanto agentes que atuam na produção das imagens técnicas contemporâneas, as IA generativas analisadas reforçam a separação entre o indivíduo real e a sua representação estereotipada, o que culmina na reprodução do racismo e seus dispositivos de manutenção, encapsulados em uma falsa objetividade.

5 CONSIDERAÇÕES FINAIS

As imagens técnicas criadas pelas IA generativas são abordadas como instrumentos de poder, pois quem controla a tecnologia, pode controlar a forma como a realidade é representada e percebida. A relação de poder aqui se manifesta nas mãos de quem modela os algoritmos e os conjuntos de dados que geram essas imagens. Uma vez que as

ferramentas de IA estão sob o controle de empresas e programadores, são agentes passíveis de replicar as estruturas racistas que acabam moldando a percepção de grupos sociais, o que evidencia uma concepção semiósica operante.

As IA generativas analisadas são parte de um novo ciclo de produção de realidades visuais que reflete, reproduz e amplifica dispositivos de racialidade preexistentes. As imagens de pessoas negras criadas por IA, não são neutras, pois podem ser impregnadas por elementos que revelam relações de poder, bem como por marcadores da branquitude e do racismo, em sua composição enquanto signos, o que torna perceptível o quanto a tecnologia se entrelaça às representações sociais e culturais em sua ação sígnica.

Ao promover uma investigação cujo foco é a relação trirrelativa entre objeto, signo e interpretante na atuação das ferramentas de IA generativa de imagens, a crítica semiótica aqui desenvolvida, pretendeu contribuir para desnaturalizar a semiose algorítmica, mostrando como as escolhas maquínicas, mesmo quando apresentadas como neutras ou objetivas, carregam consigo questões éticas e sociais que afetam a forma como raça e alteridade são percebidas.

Considera-se que as implicações deste estudo se concentram em três dimensões, que sintetizamos a seguir:

a) Dimensão ética: a necessidade de reconhecer e enfrentar o racismo algorítmico como uma questão central na pesquisa e desenvolvimento de IA a fim de garantir que as ferramentas tecnológicas não reforcem desigualdades sociais, mas atuem como instrumentos de inclusão e justiça social, sendo fundamental incorporar princípios de ética e responsabilidade social em todas as etapas de concepção e implementação das ferramentas e sistemas.

b) Dimensão operacional: implementação, por parte das empresas e instituições, de auditorias regulares nos algoritmos e modelos de treinamento a fim de identificar e mitigar vieses raciais; promover a inclusão de dados diversos e representativos na etapa de treinamento das ferramentas; desenvolver diretrizes claras e transparentes para a geração e utilização das imagens, assegurando o respeito aos princípios de diversidade e equidade.

c) Dimensão acadêmica: incentivar estudos interdisciplinares que integrem ciência da informação, semiótica, ciência da computação e estudos étnico-raciais; explorar como as dinâmicas de poder são codificadas em algoritmos e seus impactos na sociedade e, por fim, desenvolver modelos teóricos que articulem a relação entre tecnologia, representação cultural e relações de poder.

Acredita-se que estudos futuros devem ainda explorar: a) como diferentes contextos culturais e linguísticos influenciam a geração de imagens por IA; b) como as percepções de usuários oriundos de diferentes grupos raciais são moldadas nas imagens geradas por IA; c) diretrizes e *frameworks* para a criação de algoritmos mais inclusivos e sensíveis à diversidade; d) experimentos controlados para analisar o impacto das intervenções éticas em algoritmos de geração de imagens; e) outros aspectos da representação imagética, como gênero, etarismo e classe, nas análises sobre semiose e enviesamento algorítmico.

O conceito de semiose algorítmica, aqui apresentado, envolve tanto a criação de significados quanto às distorções causadas pelo enviesamento, oferece uma lente para analisar como os sistemas de IA generativa de imagem não apenas geram representações, mas também influenciam e são influenciados pelas ideologias, valores e preconceitos oriundos dos dados de treinamento. Tal abordagem crítica pode auxiliar no entendimento sobre o impacto social e cultural das tecnologias de IA e pode ser um ponto de partida para estudos e práticas que instaurem maior responsabilidade e transparência nos sistemas algorítmicos contemporâneos.

REFERÊNCIAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação**: Conceitos e Tecnologia das Máquinas de Busca. Bookman Editora, 2011.

BENGIO, Y.; LECUN, Y.; HINTON, G. Deep learning for AI. **Communications of the ACM**, v. 64, n. 7, p. 58-65, 2021.

BENTO, C. **O pacto da branquitude**. São Paulo: Companhia das Letras, 2022.148 p.
CARNEIRO, S. **Dispositivo de racialidade:** a construção do outro como não-ser como fundamento do ser. São Paulo: Editora Jandaíra. 2023.431 p.

CASTRO, L. M.; OTERO, T. P.; ROLÁN, L. X. M. **As intelixencias artificiais como xeradoras de cultura**: Exploración dos nesgos de xénero de ChatGPT. In: Epistemoloxías feministas en acción: VIII Xornada Universitaria Galega en Xénero. Servizo de Publicacións, 2023. p. 213-222.

COLLINS, P. H. **Pensamento feminista negro:** conhecimento, consciência e a política do empoderamento (D. N. Barbosa, Trad.). São Paulo: Boitempo. 2019.
CORREDERA, J. R. C. Inteligencia artificial generativa. In: **Anales de la Real academia de Doctores.** 2023. p. 475-489.

DE ZÁRATE, J. M. O. *et al.* **Sesgos algorítmicos y representación social en los modelos de lenguaje generativo (LLM)**. 2024.

FLUSSER, V. **Filosofia da caixa preta:** ensaios para uma futura filosofia da fotografia.1985. 48p.

GILLESPIE, T. A relevância dos algoritmos. **Parágrafo**, [S.l.], v. 6, n. 1, p. 95-121, jun. 2018.

GOODFELLOW, I. *et al.* Generative adversarial nets. **Advances in neural information**

HAYKIN, S. **Redes neurais:** princípios e prática. Bookman Editora, 2001.

HINTON, G. E. *et al.* **How neural networks learn from experience**. 1992.

HO, J.; JAIN, A.; ABBEEL, P. Denoising diffusion probabilistic models. **Advances in neural information processing systems**, v. 33, p. 6840-6851, 2020.

IBÁÑEZ, A. R. **Análisis de sesgos en modelos de IA generativa.** Caso de uso Inditex. 2024. Tese de Doutorado. Universitat Politècnica de València.

LUCCIONI, A. S.; AKIKI, C.; MITCHELL, M.; JERNITE, Y. **Stable bias:** evaluating societal representations in diffusion models. 37th Conference on Neural Information Processing Systems. 2023. p. 1-39.

MANINI, M. P. Análise documentária de imagens. **Informação & Sociedade:** Estudos, v. 11 n.1 2001, n. 1, 2001.

MCCULLOCH, W.S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, v. 5, p. 115-133, 1943.

MOURA, C. M. S.; BRAGA, T. E. N. A Inteligência Artificial e a criação de conteúdo: os vieses que habitam entre nós. In: Anais do Workshop de Informação, Dados e Tecnologia-WIDaT. 2023.

PEIRCE, C. S. **Collected Papers of Charles Sanders Peirce**. Versão eletrônica.

PEÑARANDA, S. G. **Ética e Inteligencia Artificial**. 2023.

PINTO, J. **1,2,3 da semiótica**. Belo Horizonte: Editora UFMG, 1995. **processing systems**, v. 27, 2014.

RADFORD, A. et al. Learning transferable visual models from natural language supervision. In: International conference on machine learning. PmLR, 2021. p. 8748-8763.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by backpropagating errors. **Nature**, v. 323, n. 6088, p. 533-536, 1986.

SHATFORD, S. Some issues in the indexing of images. **Journal of the American Society for Information Science**. [Washington, USA], v. 45, n. 8, p. 583-588, Set. 1994.

SILVA, T. **Racismo algorítmo:** inteligência artificial e discriminação nas redes digitais. São Paulo: Edições SESC, 2022.

SILVA, T. Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código. In: **Anais** do VI Simpósio Internacional LAVITS. Salvador, Bahia, Brasil. 2019.

SMIT, J. W. A representação da imagem. **Informare** – Cadernos da Pós Graduação, Ci. Inf., Rio de Janeiro, v.2, n.2, p. 28-36, jul./dez. 1996.

VASWANI, A. *et. al.* Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

NOTAS

AGRADECIMENTOS

CONTRIBUIÇÃO DE AUTORIA
Concepção e elaboração do manuscrito: J. Assis, M. A. Moura
Coleta de dados: J. Assis
Análise de dados: J. Assis, M. A. Moura
Discussão dos resultados: J. Assis, M. A. Moura
Revisão e aprovação: J. Assis, M. A. Moura

CONJUNTO DE DADOS DE PESQUISA
FINANCIAMENTO
Não se aplica.

CONSENTIMENTO DE USO DE IMAGEM
Não se aplica.

APROVAÇÃO DE COMITÊ DE ÉTICA EM PESQUISA
Não se aplica.

CONFLITO DE INTERESSES
Não se aplica.

PUBLISHER
Universidade Federal de Santa Catarina. Programa de Pós-graduação em Ciência da Informação. Publicação no Portal de Periódicos UFSC. As ideias expressadas neste artigo são de responsabilidade de seus autores, não representando, necessariamente, a opinião dos editores ou da universidade.

EDITORES
Edgar Bisset Alvarez, Ana Clara Cândido, Patrícia Neubert, Genilson Geraldo, Jônatas Edison da Silva, Mayara Madeira Trevisol, Edna Karina da Silva Lira e Luan Soares Silva.

HISTÓRICO
Recebido em: 15-10-2024 – Aprovado em: 31-01-2025 – Publicado em: 14-03-2025