Imagens fragilizadas e preconceito algorítmico latente: estudo de caso de imagens geradas por IAG[1]

Fernanda Sevarolli Creston Faria Kistemann [2]

Resumo

O presente estudo demonstra como duas IAG se comportam na criação/produção de imagens e "transformação" de fotos de modo a perpetuar o preconceito algorítmico e microagressões evidenciadas através da observação de foto ou imagem-origem e foto ou imagem-resultado. Além disso, a violência demarcada pela colonização de dados influencia fortemente o aparato digital e permite a luta desigual pelo poder ainda nas mãos do branco heterossexual e europeu. Silva (2020, 2025), Vieira (2025), entre outros autores, colaboram na compreensão deste estado de alerta que se apresenta no mundo digital e que deve ser combatido, pois se trata de formas agressivas de preconceito e proliferação de violência.

Palavras-chave: Preconceito; microagressões; Inteligência Artificial Generativa (IAG); imagens.

Introdução: um ponto de partida

A tecnologia amparada por Inteligência Artificial (IA) tem evoluído rapidamente. A IA Generativa (IAG) é uma das tecnologias mais procuradas e apresenta resultados surpreendentes.

Apesar deste tipo de tecnologia ser uma ameaça altamente periculosa para os recursos natureza, pois consome grande volume de energia para sua manutenção e existência, a procura por tecnologias baseadas em IA cresce de forma descontrolada.

As ferramentas que utilizam a IA para obter resultados em diversos campos do conhecimento estão tomando espaço na vida cotidiana de forma contundente. Pesquisas

[1] Trabalho apresentado no ET 2. Decolonização, algoritimização e racismo do XVIII Simpósio Nacional da ABCiber – Associação Brasileira de Pesquisadores em Cibercultura. Realização Faculdade Cásper Líbero, nos dias 11 a 13 de novembro de 2025.

[2] Doutoranda do Programa de Pós-graduação, UFJF, Grupo de Pesquisa Namidia, Bolsa Capes, fernandasevarolli@gmail.com.

acadêmicas e escolares, buscas administrativas e jurídicas, ampliação de conhecimentos médicos, etc.

Mas os resultados fornecidos por este tipo de ferramenta são confiáveis?

Qual a repercussão da IAG no dia a dia das pessoas?

Há isenção nos resultados ou existe algum traço de violência ou preconceito?

Estas e outras perguntas ecoam na sociedade de maneira cada vez mais comum, já que o uso da IA e IAG é crescente. A maneira como se usa a ambas é uma preocupação importante a ser considerada em um cenário de polarização e desinformação. Além do exposto, a necessidade de regulação, governança, entre outros fatores que permeiam a presença e exploração da tecnologia e da virtualidade no meio social é uma reflexão que busca modos de proteger os usuários contra possíveis irregularidades e agressões digitais.

Há muitos impasses que suscitam um "olhar desconfiado" em relação à utilização de IA, principalmente a do tipo generativa. Erros de referência, atribuições falsas, etc., são alguns elementos que fazem do digital "um local" de desconfiança que merece atenção.

Apesar do crescente uso da referida tecnologia, pesquisa acadêmicas preocupadas com os resultados obtidos tem se fortalecido. Esforços científicos a respeito afirmam que, ao utilizar a IA, "o resultado precisará ser cuidadosamente verificado, já que o software parece gerar conteúdo impreciso, baseado em fontes de ideias relatadas de forma imprecisa" (Dwivedi *et al.*, 2023, p. 5). Apesar de ser deveras necessária, a assertiva do autor não se traduz na realidade de utilização da IA, já que a não verificação de dados continua sendo um motor de propagação de *fake news*, desinformação e potentes "construções" baseadas em IA distantes da realidade.

Quando se menciona a função generativa deste tipo de inteligência, rememora-se exemplos que podem ilustrar as reflexões ora expostas. Em 2020, época em que se passava pela pandemia de COVID-19[3], a Fiat lançou o modelo comemorativo Fiat Strada. A peça audiovisual utilizada como publicidade trouxe Elvis Presley, criado em *deep fake*, para dar vida a uma personagem lendária no mundo do rock. Ele dirigia o novo modelo da Fiat para demonstrar que ambos são lendas[4] até os dias de hoje.

---

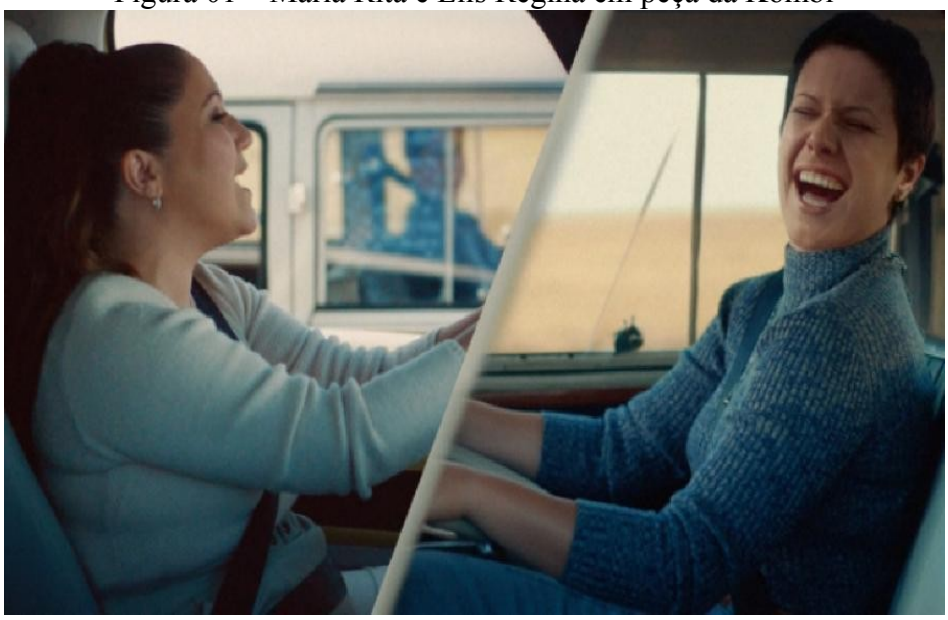
[3]

[4] Ao clicar na palavra "lendas", é possível assistir a referida peça audiovisual na íntegra em um canal particular do *Youtube*.

Como já mencionado, passava-se por uma epidemia de proporções mundiais e, ademais, questões que envolviam a ética no uso da IAG não estavam em discussão dada à crise sanitária que se vivenciava globalmente, que era algo de maior importância naquele momento. Portanto, não havia engajamento suficiente que suscitasse uma discussão de grande porte sobre o assunto. Considera-se também que a IA não era tão potente à época e não tinha grande penetração e divulgação no meio social, o que corroborou com a não discussão do assunto.

Após o término da pandemia, em julho de 2023, uma nova criação gerou pertinente discussão de cunho ético no campo da Comunicação. Foi uma discussão que envolvia grande polêmica a partir do lançamento de uma campanha publicitária que "colocou" (ressuscitou) Elis Regina e sua filha, Maria Rita, dirigindo lado a lado, em um mesmo vídeo, conforme é possível observar na Figura 01.

Figura 01 – Maria Rita e Elis Regina em peça da Kombi



Fonte: Uol (2024).

A mencionada peça audiovisual foi resultado do uso de IAG para "trazer à vida" a cantora Elis (1945-1982). De modo mais aperfeiçoado do que o vídeo de 2020 de Elvis Presley, a cantora está na tela dirigindo uma kombi furgão modelo antigo. Na mesma estrada, ao seu lado, aparece

sua filha, Maria Rita, também cantora, dirigindo uma ID Buzz (kombi elétrica). Mãe e filha cantam o sucesso "Como nossos pais".

O lançamento deu potência a discussões já em curso sobre o uso de IA e, inclusive,

> O novo comercial da Volkswagen, que usou inteligência artificial para unir Elis Regina e Maria Rita ao som de "Como Nossos Pais"' gerou uma nova discussão na web e dividiu opiniões. (...) Internautas abordaram um possível perigo do uso da inteligência artificial e da *deepfake* em um futuro próximo, e o fato de recriarem pessoas que já morreram, por meio desses sistemas. Por outro lado, algumas pessoas ficaram impressionadas com o poder da IA e das coisas que só ela pode fazer (Pazero, 2023, *online*).

Além dos internautas, "o gerente de projetos da Safernet, Guilherme Alves, que atua na defesa dos direitos humanos na internet, disse que o comercial nos leva a refletir também sobre os perigos causados pelo avanço desse tipo de tecnologia" (Pazero, 2023, *online*).

> "Os usos são diversos e preocupantes porque o potencial de desinformação da *deepfake* é gigantesco e fica cada vez mais explícito. Claro que um comercial desse é emocionante. Mas é preciso pensar que existem riscos muito grandes para debate público e informação das pessoas. Aqui a questão que pega é que não temos uma regulação sobre o uso da *deepfake* […] uma regulação específica sobre o que pode ou não ser feita, tipos de cuidados quando falamos dessas tecnologias, e isso é urgente. As pessoas tendem a acreditar mais no vídeo do que texto, esse potencial é bastante perigoso e deve ser presente quando discutimos os tipos de regulação dessas tecnologias", esclareceu o profissional (Pa zero, 2023, *online*).

Apesar da tentativa de surpreender o público com a novidade e reiterar o conceito do velho e do novo presentes no enredo da peça audiovisual, a Conar (Conselho Nacional de Auto-Regulação Publicitária) interviu e a Volkswagen retirou a peça publicitária de seus meios oficiais de comunicação (TV, redes sociais e internet em geral). A empresa foi processada por questões de direitos autorais e falta de ética no uso da imagem de uma pessoa falecida, já que Elis não poderia negar ou aceitar sua aparição no vídeo por estar morta.

O vídeo foi deletado da plataforma oficial da empresa no *Youtube;* contudo, fãs já haviam hackeado o vídeo da plataforma oficial e disseminado seu conteúdo por diversos outros sites[5].

[5] Ao clicar na palavra "sites", é possível assistir a referida peça audiovisual na íntegra em um canal particular do *Youtube*.

São dois exemplos significativos com a presença de ícones da música. A cantora Elis Regina é representante da música popular brasileira (MPB) e Elvis Presley é um ícone da música pop mundial. Ambos são deveras famosos no meio musical e sua presença ainda é significativa em vários dispositivos de comunicação.

São exemplos polêmicos, mas não são os únicos existentes em relação à IAG. Apesar do exposto e de outros impasses relativos ao seu uso de forma ética, a tecnologia que desenvolve a IA continua avançando com inovações mais determinantes e com mais desafios ligadas à internet a cada dia. Normatização, controle, vigilância e segurança são alguns elementos necessários e em discussão sobre o assunto.

Outra situação que preocupa em relação à utilização da IA, principalmente a IAG, é a presença de preconceito algorítmico e de microagressões nos resultados obtidos através de ferramentas com a tecnologia mencionada.

Estes dois conceitos (preconceito algorítmico e microagressões) merecem destaque no presente texto, que busca demonstrar como eles acontecem através do estudo de caso de conteúdos resultantes de IAG. Antes do referido estudo, revisa-se alguns teóricos e suas concepções sobre os dois conceitos a seguir.

Aprofundamento teórico: quem está refletindo sobre o assunto?

Além das questões éticas e morais que envolvem a utilização da IA e a IAG, o preconceito algorítmico e as microagressões também ocorrem e corroboram com o clima de preocupação sobre tais tecnologias. Estes aspectos estão presentes em vários exemplos de utilização de ferramentas digitais baseadas em IA, exemplos estes que serão discutidos posteriormente neste texto.

Ressalta-se que os aplicativos que utilizam IA e IAG funcionam a partir de códigos e algoritmos fornecidos para seu funcionamento ainda no período de desenvolvimento da tecnologia de ação empregada em aplicativos diversos. Tais códigos e algoritmos utilizados e executados são, não raro, criados a partir de um imaginário branco, masculino, heterossexual e europeu, conforme aponta Lugones (2014, p. 936).

Neste caso, não se leva em consideração outros aspectos que não aqueles inseridos na tecnologia e mencionados por Lugones. Por consequência, os resultados encontrados a partir da utilização das tecnologias em questão atendem às características dos dados de controle que engendram seus sistemas. Lembrando que os dados de controle também são inseridos durante o desenvolvimento dos aplicativos para colaborar na melhor execução de suas tarefas.

Infere-se, pois, que as características referentes a negros, mulheres, homossexuais, travestis, pessoas trans, etc., ou seja, algumas (ou várias) minorias são ignoradas. Mas ao existir uma situação de ignorância de dados, infere-se também que o preconceito se faz presente, oportunizando situações de ridicularização e violência contra minorias a partir dos resultados obtidos na utilização de IAG.

Lugones (2014) menciona que o "homem europeu, burguês, colonial moderno tornou-se um sujeito/agente, apto a decidir, para a vida pública e o governo, um ser de civilização, heterossexual, cristão, um ser de mente e razão".

Assim, depreende-se que a autora se aproxima da lógica de criação de códigos digitais que priorizam como "melhor" aqueles indivíduos com as características destacadas por ela. Os demais indivíduos, ou seja, os "outros", as minorias sem as características citadas principalmente, estariam à margem da realidade e poderiam ser considerados "ruins" para sua utilização como parâmetro no âmbito generativo, no caso da IAG.

Contudo, desmerecer as minorias não aceitando suas características e elegendo um modelo único de referência, como o citado por Lugones, incita microagressões cada vez mais crescentes em plataformas digitais. Silva (2019, 2020) demonstra, em algumas de suas pesquisas sobre o assunto, que o nome "microagressão" parece minimizar algo que, ao contrário, deveria ser alardeado.

Para compreender o impacto das microagressões, no sistema digital ou fora dele, é preciso saber o que significa uma microagressão e como ela ocorre. Trata-se de palavras e/ou situações constantemente utilizadas para minimizar uma agressão para que ela não seja evidenciada, mas sim apagada das relações sociais e/ou digitais. Por exemplo:

(1) Quando alguém agride verbalmente uma pessoa preta, para se assegurar que não será punida ou exposta, diz "Foi só uma piada!". Mas se algo foi preferido como piada, sabendo que a maioria das piadas têm cunho pejorativo, prioritariamente quando são utilizadas contra minorias, tem o intuito explícito de ridicularizar, ofender, agredir.

(2) Esta e outras situações podem e devem ser complexificadas. Elas comumente ocorrem e, ainda, suscitam o seguinte comentário: "Ah, foi brincadeira!". Observe que se foi dito é porque há o objetivo de criticar alguém de alguma forma e o referido comentário ocorre como forma amenizar a situação. Perceba que já não se pode chamar de brincadeira o que se apresenta como uma violência simbólica. Tal violência é cometida a partir da fala preconceituosa e agressiva de alguém direcionada a outrem.

Estas e outras palavras e expressões tem o intuito de ridicularizar as minorias, como já mencionado, e se passar por algo a não ser salientado, pois quem comete este tipo de ato, quer que o mesmo fique despercebido, ignorado. Na verdade, este tipo de palavra, expressão e/ou situação evidencia a presença de preconceito e de um racismo latente na sociedade e na *wide world web* (www) de forma brutal.

Outra realidade, já mencionada, e que ocorre no meio digital, está ligada ao preconceito algorítmico, especificamente no uso de IAG. Segundo Vieira (2024),

> (...) os sistemas algorítmicos baseados em IA Generativa são construídos utilizando o que chamamos de Redes Neurais Adversárias (GANS). Esse tipo de rede neural tem como objetivo o aprendizado e geração de novos dados em diferentes formatos, como: textos, imagens, áudios, vídeos etc. (Vieira, 2024, p. 10).

Ao compreender-se, em suma, a IAG, há a percepção de que se trata de uma ferramenta criada para agrupar informações e criar imagens (textos, etc.) de acordo com o que ela recebe e condiciona, seguindo índices iniciais de programação e ajustes de aprendizado sistêmico (de acordo com as imagens, textos, etc. de controle inseridos para início de suas projeções e resultados).

Estes objetos de controle acumulados pela IAG, juntamente com as informações recebidas ao longo de suas utilizações, vão atuar diretamente nos resultados por ela emitidos, pois eles formam uma espécie de banco de dados referencial, além do aporte existente na rede da internet. Portanto, ao se ponderar sobre a IAG e os resultados obtidos através dela nos últimos anos, encontra-se o preconceito refletido em muitas imagens (textos, etc.) por ela criadas. Tais resultados denotam a utilização de parâmetros que coadunam com o "branqueamento" no desejo do que vai se obter como resultado em detrimento de outras etnias, *a priori*, tendo como referência o que se chama de preconceito algorítmico. Observe que Vieira (2024) relata exemplos deste tipo de preconceito, já que

> (...) nos últimos anos, relatos de usuários do mundo todo têm revelado as fragilidades dos modelos de IA Generativa (como ChatGPT, Midjourney, DALL-E entre outros). Uma matéria e pesquisa realizada pela Bloomberg, mostrou os resultados de geração de imagens de trabalhadores para 14 diferentes empregos (classificados de "alta remuneração" ou "baixa remuneração" nos EUA). Os resultados relevam o racismo algorítmico enraizado nas tecnologias de forma que o modelo chegou a gerar imagens de pessoas com tons de pele mais escuros 70% das vezes para a palavra-chave "trabalhador de fastfood", embora 70% dos trabalhadores de fast-food nos EUA sejam brancos (Vieira, 2024, p. 10-11).

Ao lidar com plataformas digitais, algoritmos, códigos e sistemas digitais é possível observar o quão sério é lidar com sistemas estruturados em relação à sociedade em si, pois cada qual evidencia uma realidade ilimitada dentro de sua rede constitutiva e da rede da internet que, além de sua característica de infinitas conexões e nós, ainda tem uma face oculta: a possibilidade de esconder seus reais autores.

De acordo com Barros (2025, p. 4), por se tratar de uma tecnologia interseccional, a IAG "(...) segue padrões de violências identificadas há muitas décadas por pesquisadoras negras e que nestes códigos não existem acidentes ou neutralidades, mas é sim o que Lélia Gonzalez chama de uma neurose cultural".

Quando Barros (2025) cita Lélia Gonzalez, ela procura dar suporte à ideia de descontrole inserida nos algoritmos e dos resultados encontrados a partir deles. Não se trata de chamar aqueles que pesquisam o assunto como neuróticos, mas sim de visualizar a situação como uma neuropatia cultural, ou seja, como uma doença. Por doença, compreende-se aquilo que deve ser tratado para

ter cura; portanto, a neurose cultural mencionada por Barros é uma doença que precisa ser sanada no âmbito digital e social em geral.

Diante do exposto, apresenta-se o estudo de caso sobre a utilização de duas ferramentas baseadas em IAG para demonstrar como as microagressões e o preconceito algorítmico, mencionados anteriormente, ocorrem nos resultados de dois aplicativos como forma de reflexão e ação contra os danos que podem ocasionar tais resultados no meio social.

Metodologia de pesquisa

Utilizamos a pesquisa qualitativa exploratória através da metodologia de análise de conteúdo (Bardin, 1977) e, como procedimento metodológico, o estudo de caso (Yin, 2015). Os dados foram colhidos do arquivo pessoal da autora a partir de testagens em duas possibilidades de uso de IAG (dois aplicativos), nas quais foram detectados comportamentos abusivos e que indicam a presença de preconceito algorítmico e microagressões.

A análise dos dados levou em consideração a observação do poder simbólico (Bourdieu, 1998) daqueles que supostamente seriam os indivíduos cujas características são consideradas parâmetros para a obtenção de resultados em IAG e da disputa de poder (Foucault, 2003), entre minorias e os considerados “melhores”.

Tais aspectos estão comumente presentes em ambientes digitais, conforme será possível visualizar nos exemplos analisados. Já as disputas de poder foram analisadas a partir da observação do fenômeno do colonialismo de dados, segundo Couldry e Mejias (2018), como ferramenta de motivação do preconceito algorítmico e das microagressões empreendidas em variadas situações, bem como comportamentos que corroboram com tais situações.

Objetos de estudo e análise

Em 2024, houve um *boom* no aperfeiçoamento e uso de aplicativos de IAG e, inclusive, ocorreu a distribuição e acesso gratuito a vários deles. Neste ínterim, alguns técnicos exploraram, à época, a alternativa de inserir (apenas) o rosto de pessoas nas mais diversas situações e corpos.

Neste sentido, a primeira observação se deu através dos resultados obtidos no aplicativo "*Toonme*". Trata-se de um aplicativo que utiliza imagens pré-elaboradas para a inserção do rosto de quaisquer pessoas em diversas situações e realidades. Podem ser recriadas imagens de uma pessoa pilotando uma motocicleta; uma rainha altiva em seu trono; um príncipe em um cavalo branco, etc. Reiterando, percebe-se que é um aplicativo com possibilidade de inserção de qualquer rosto em situações que resultarão em animações ou, em inglês, *cartoons*, já que as imagens-resultado são/serão alegóricas.

O aplicativo pode ser encontrado em *site* do mesmo nome ou na *app store* do Google, conforme a publicidade da Figura 02.

Figura 02 – Publicidade/Ícone *Toonme*

Fonte: *Toonme* (2025).

A pesquisadora se encontrava no início de seu doutorado e julgou pertinente testar o aplicativo, pois em um primeiro momento de sua pesquisa era pertinente refletir sobre as tecnologias e seu uso nas Humanidade Digitais[6]. Além de refletir sobre as tecnologias, a pesquisadora trabalha com o fantástico, tendo nos resultados em forma de animação uma forma de

[6] Em suma e de forma bastante simplificada, o termo se refere ao uso de tecnologias dentro da área de Ciências Humanas, como forma de otimização de dados e resultados.

aporte visual para suas considerações futuras. Diante disso, foi colocado o rosto dela em princesas de contos de fadas (baseadas nas princesas da Disney[7]).

Enquanto pessoa parda, a pesquisadora esperava que, ao solicitar que seu rosto fosse colocado no corpo de uma princesa, o aplicativo *Toonme* pudesse trabalhar com o contexto de sua aparência e transformá-la como um todo, ou seja, mantendo a cor parda, conforme demonstra o exemplo do ícone na Figura 02, o qual manteve as características originais na imagem-resultado.

Desta experimentação foi colhida a primeira figura a ser analisada.

Testagem 1 – Apresentação e análise de figuras: Cadê meus traços que estavam aqui?

Figura 03 – Princesa Bela

Fonte: Arquivo pessoal da autora (2025).

[7] É importante mencionar que o referencial visual de princesas utilizado no aplicativo se remete às produções da Disney, já que outros modelos e visuais de realeza não são considerados no aplicativo, pelo menos no momento de sua utilização.

A Figura 03 apresenta o resultado da solicitação feita ao *Toonme*, qual seja a criação da Princesa de "A Bela e a Fera" como foto-guia da pesquisadora. O resultado expôs uma mulher branca, e não parda. O rosto da pesquisadora foi utilizado como base, mas a cor de sua pele foi alterada, o cabelo foi alisado e, ainda, ocorreu a mudança de tonalidade de seu cabelo e o formato do rosto foi afinado.

A IAG foi novamente interpelada, agora para colocar o rosto da pesquisadora no corpo de outra princesa: "Branca de Neve". Apesar de o nome sugerir a cor da pele da princesa de antemão, o teste era necessário para verificar o comportamento do aplicativo a partir da mudança de personagem.

Figura 04 – Branca de Neve

Fonte: Arquivo pessoal da autora (2025).

Na Figura 04, novamente, a cor da pele, o formato do rosto, o cabelo e, agora também, o corpo da pesquisadora foram alterados seguindo padrões europeus de beleza e aparência, ou seja,

ocorreu o "branqueamento" generalizado da imagem-origem na obtenção do resultado imagem-resultado.

Observa-se, pois, que o aplicativo utilizado lida com um "sistema de práticas contra grupos racializados que privilegiam e mantêm poder político, cultural e econômico para os brancos no espaço digital" (TYNES *et al*, 2019, p.195). A citação de Tynes *et al* corrobora com a ideia de violência simbólica, analisada a partir da visão de simbólico de Bourdieu (1998) e, ainda, consolida os padrões de poder expostos por Foucault (2003), demonstrando que o privilégio branco sobre os demais grupos e etnias prevalece, inclusive, no âmbito digital.

Ademais, nos casos das Figuras 03 e 04, foi observado o preconceito algorítmico alterando os caracteres reais da pesquisadora, que deveriam ser reconhecidos em foto enviada à IAG para a criação/produção da princesa; contudo, foram utilizados padrões e parâmetros previamente organizados que, além de preconceituosos, podem ser vistos como uma espécie de microagressão ao se ignorar a imagem original e resultando no "branqueamento" não esperado/almejado.

Este "branqueamento", tido como proposital, incomodou e levou à observação do comportamento de outra IAG em relação ao uso e produção/criação de imagens. Conforme os resultados obtidos através do aplicativo *Toonme* foram observados os seguintes elementos para a verificação dos resultados em nova IAG: cor da pele, gênero e raça, prioritariamente, já que estes elementos foram os mais afetados na experimentação do referido aplicativo utilizado anteriormente. O olhar não seria apenas sobre o resultado imagético. Agora foi necessário realizar também uma reflexão amparada teoricamente nos novos resultados criados/produzidos.

A seguir, apresenta-se a comparação entre a imagem fornecida pelo "*Toonme*", em 2024, e a imagem fornecida pelo "ChaGpt", em 2025. No ChatGPT, observa-se um comportamento gordofóbico e também o exagero nos traços latinos (Quijano, 2005, p. 117).

Figura 05 – *Toonme* (2024)

Figura 06 – ChaGpt (2025)

Fonte: Arquivo pessoal da autora (2025).

Para a construção da primeira imagem (Figura 06 – imagem-resultado) foi utilizada como referência a princesa Mérida (conforme criação da Disney, Figura 05). A tentativa do *Toonme,* Figura 05, foi a de se aproximar o máximo possível da princesa original, mesmo que os traços da imagem-guia não correspondessem ao que fora apresentado. A Mérida da animação é branca, com cabelos encaracolados e ruivos. A pele, o rosto e a corpo foram alterados mais uma vez.

Já o ChatGPT, Figura 06, transformou a pesquisadora, uma mulher parda, de cabelos escuros e acima do peso, mas não muito obesa, em uma figura latina, com traços exagerados que marcam a latinidade de forma pejorativa, tingiram sua pele em um tom de amarelado, justificado talvez por um filtro sépia que toma toda a imagem. Observe que os óculos que a pesquisador usa, neste caso, foram mantidos.

Constata-se, após observação cuidadosa das Figuras 05 e 06, que as microagressões e o preconceito algorítmico estão explícitos e demonstrados nos resultados obtidos nestas testagens. Percebe-se que, em relação ao preconceito e às microagressões, “estão relacionados a grandes

fragilidades existentes dentro de sistemas baseados em IA Generativa: (i) as bases de dados de treinamento, (ii) opacidade e (iii) veracidade de conteúdo" (Vieira, 2025).

Vieira (2025) ainda ressalta que

> No caso da IA Generativa, os vieses são em grande parte inseridos justamente através dos dados de treinamento. Como a IA Generativa trabalha sob a premissa de encontrar padrões nos dados, esta pode incorporar estereótipos de gênero e preconceitos existentes nos dados de treinamento. Por exemplo, como o treinamento inclui textos e imagens que retratam pessoas negras em papéis estereotipados, os modelos continuam reforçando essas representações nos conteúdos gerados. Dessa forma, um viés histórico se faz muito presente (Vieira, 2025, p.12).

Silva (2020, p.431) defende que "os estudos sobre a branquitude são uma chave importante para entender os modos pelos quais as tecnologias automatizadas demonstram continuamente vieses racistas". O autor cita Bento, que define branquitude como "um lugar de privilégio racial, econômico e político, no qual a racialidade, não nomeada como tal, carregada de valores, de experiências, de identificações afetivas, acaba por definir a sociedade. Branquitude como preservação de hierarquias raciais, como pacto entre iguais (Bento, 2002, p.7). O autor relata ainda que

> A manutenção e reprodução dos privilégios da branquitude partindo de uma centralidade evocativa à Europa se ligaram histórica e economicamente à dominação colonial e neocolonial, com desdobramentos da ciência à tecnologia, mas sempre através da evitação ao debate sobre raça. Mills chama este esforço coletivo de inverter a epistemologia de uma "epistemologia da ignorância", um "padrão particular de disfunções globais e locais (que são funcionais psicológica e socialmente), produzindo o resultado irônico que os brancos em geral serão incapazes de entender o mundo que eles próprios formataram" (MILLS, pos.327, trad. nossa).

Sobre a opacidade, Silva (2020, p. 432) propõe a definição de que é "o modo pelo qual os discursos hegemônicos invisibilizam tanto os aspectos sociais da tecnologia quanto os debates sobre a primazia de questões raciais nas diversas esferas da sociedade – incluindo a tecnologia, recursivamente", gerando, assim, uma dupla opacidade na visão do autor.

O indivíduo que utiliza a IAG vê sua identidade alterada ou grosseiramente acentuada por traços que ela (a IAG) julga que sejam "melhores" que aqueles presentes na imagem-guia, pois

assim foram inseridos ou "aprendidos" pela tecnologia implantada na aplicativo em uso através de aprendizagem de máquina.

Testagem 2 – Apresentação e análise de fotos: "Acho que eu vi um gatinho!"

Outro experimento, também em formato de testagem, foi realizado a partir da provocação de um *post* encontrado na plataforma da rede social Instagram. O *post*, da referida plataforma, apresentava o relato de uma jovem que utilizou uma IAG, sem mencionar qual, para verificar como seriam seus gatos se fossem crianças, seres humanos.

A proposta é interessante ao se imaginar que os gatos tem cores variadas, então, surge a seguinte questão e curiosidade por parte da pesquisadora: como seria a interpretação da IAG das variadas cores dos gatos?

A experimentação utilizou o ChatGPT e fotos de gatos da pesquisadora a partir do questionamento de como eles seriam se fossem crianças, seres humanos. Para tanto, foi construído um *prompt*[8] solicitando que fosse transformado o gato ou a gata na foto inserida na tecnologia em menino ou menina.

A partir do *prompt criado, utilizamos* seis fotos de seis gatos de cores diferentes, uma de cada vez para respeitar o gênero de cada um deles e, desta forma, ter dados para a análise dos resultados. Salienta-se que foi ignorada e não incluída no *prompt* a idade e/ou outras informações para a confecção da foto-resultado, uma vez que o interesse na geração das imagens se prendia à cor dos animais.

A seguir, os resultados obtidos.

Foto 01 – Gato preto e branco

Foto 02 – Menino branco

[8] *Prompt* utilizado para a obtenção das fotos foi o seguinte: para meninos – "transforme meu gato em um menino" e, para meninas, "transforme minha gata em menina".

Fonte: Arquivo pessoal da autora (2025).

Nas Fotos 01 e 02, encontra-se a foto-guia e a foto-resultado: um gato preto e branco, com predominância da cor branca, resultou em um menino branco respectivamente.

Foto 03 – Gata laranja

Foto 04 – Menina branca

Fonte: Arquivo pessoal da autora (2025).

As Fotos 03 e 04, apresentam, respectivamente, uma gatinha branca e laranja na foto-guia, a qual resultou em uma menina branca (cabelos ruivos) na foto-resultado.

Contudo, sobre dois gatinhos e os resultados obtidos a partir de suas fotos, os quais a IAG "julgou serem diferentes", quando foram transformados em crianças, não ficaram com suas cores de origem. Observe:

Foto 05 – Gato preto

Foto 06 – Menino branco com "blusa preta"

Fonte: Arquivo pessoal da autora (2025).

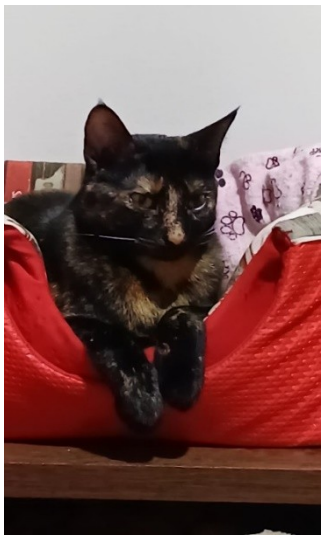
O ChatGPt se comportou aqui da mesma forma que o aplicativo "*Toonme*", nas figuras analisadas anteriormente. Ele tomou a Foto 05 (foto-guia) de um gato preto e o transformou em um menino branco, conforme a Foto 06 (foto-resultado). Inferi-se, pois, que a IA "tentou" minimizar o problema, para evitar uma interpetação preconceituosa, vestindo o menino branco com uma blusa preta, ou seja, foi uma tentativa "exdrúxula"para que o menino (foto-resultado) não ficasse "tão diferente" do gato preto (foto-guia).

Seria uma ironia se não se estivesse falando de resultados obtidos a partir do uso de aplicativos que utilizam a IAG para geração de imagens (neste caso).

Não houve ironia ao se colocar no garoto branco uma blusa preta. Os desenvolvedores, programadores e reguladores do sistema que desenvolveram o algorítimo do ChatGPt ignoraram a cor do gato, só se atentaram que era um gato macho e que deveria ser transformado em um menino. A predonminância da cor preta do animal não foi interpretada pela IA como a cor que deveria ser da pele no menino. O algoritmo apenas sugeriu uma vestimenta preta para "sanar este problema", e não interpreto que a cor do gato deveria, portanto, ser a cor da pele do menino.

Foto 07 – Gata mesclada

Foto 08 – Menina com "blusa mesclada"



Fonte: Arquivo pessoal da autora (2025).

Na Foto 07 (foto-guia), apresenta-se uma gatinha mesclada que poderia ter sido "transformada" em uma menina parda ou até mesmo morena. Contudo, o resultado que o ChatGPt entregou na Foto 08 (foto-resultado) foi uma menina branca com sardas e, ainda, vestindo uma "blusa mesclada".

Novamente é usada uma estratégia para minimizar a situação relativa à cor da pele, mesmo comportamento verificado na Foto 06 (foto-resultado).

Trata-se de uma microagressão evidenciada de forma conjunta com o preconceito algorítimico, pois, como mencionado anteriormente, este tipo de dado foi inserido na ferramenta por seres humanos.

Percebam que tanto o preconceito, quanto a sutileza (ou não), seguida de ironia, não são características cibernéticas. São marcas humanas ensinadas e/ou transmitidas à IAG através da inserção de aprendizagem "maquínica", códigos de sistema internos e algoritmos automatizados para tanto. A inserção de dados e códigos que incitam o preconceito e as microagressões, conforme observados nos resultados fornecidos pelo ChatGPT, são realizados por algorítmicos desenvolvidos a partir de interesses humanos.

As características que nos surpreenderam pela mudança de identidade visual tanto no caso das imagens de princesas, quanto na "transformação" dos gatinhos são inseridas nos aplicativos a partir de códigos e algoritmos que, posteriormente, são implantados em plataformas, IAG, etc., de

possíveis mecanismos generativos para reprodução em seus sistemas de características selecionadas de acordo com padrões e parâmetros previamente escolhidos. Ou seja, o uso e/ou inserção de dados para utilização é realizado e reproduzido a partir do preconceito e da violência iniciados na mente humana, e não no algoritmo.

Palavras finais...

A chamada inteligência artificial generativa (IAG), popularizada recentemente através de serviços como ChatGPT, Dall-E, Bard, Claude e outros tantos, produziu deslumbres sobretudo em torno da aparente possibilidade de produzir conteúdo de forma autônoma e criativa.

Refletindo sobre os processos de desenvolvimento destas tecnologias, porém, artistas e ativistas têm denunciado o caráter extrativo das empresas privadas que se apropriaram de criatividade e conhecimentos coletivos para fins privados.

A disponibilidade em escalada de serviços *freemium* multiplicou os casos de representações nocivas, desinformação e reprodução de pseudociência pelos serviços de IA. Adicionalmente, a voracidade energética e predatória da infraestrutura necessária para os serviços gera impactos ambientais descomunais ainda em mensuração**.**

Sobre o exposto, o presente estudo apresenta evidências de resultados ofensivos obtidos por dois aplicativos, selecionados para este trabalho, que utilizam a IAG para a geração de resultados.

Os resultados apresentam, com clareza e riqueza de detalhes, que as tecnologias ora utilizadas se comportam a partir de vieses de preconceito e violência, evidenciado o que Silva (2020, 2025), Barros (2025), entre outros autores, têm analisado como preconceito algorítmico e microagressões.

Observou-se ainda, a partir dos resultados apresentados, que não há mágica ou ironia nos contextos estudados e apresentados. Há sim seres humanos que colocam suas identidades, preconceitos e discursos nos sistemas a serem utilizados por outrem. As agressões sofridas, repetidas e marcadas nos contextos virtuais e digitais não podem ser observadas como brincadeiras

e/ou piadas, pois sãos marcas profundas sociais antigas da violência que ainda circunda a espécie humana.

A simbologia impetrada nos comentários ou na atuação humana através dos algoritmos que motivam microagressões no ambiente digital são dolosas e causam danos reais às pessoas, seja no âmbito material, pessoal e/ou emocional. O que parece brincadeira tem forte cunho simbólico-ideológico destilado na vida de pessoas de pele retinta (Bourdieu, 1998), principalmente.

Ademais, a colonização de dados se reflete no meio digital como uma forma de colonização de pessoas e de hábitos que enraízam a conduta escravocrata que os brancos conduziram por séculos, vinda de um passado sangrento e latente.

Além da marca de raça, há também a marca de gênero, de cor, de classe, de minorias, enfim, marcas humanas de preconceito e violências diversas que continuam vivas e reificadas nos ambientes virtuais, alimentando imaginários de ódio, preconceito e aversão (Maffesoli, 1988, 1996) que não deveriam ou não poderiam mais existir ou serem compartilhados como se fossem comuns e aceitáveis.

O branqueamento forçado, a gordofobia em forma de humor, a latinidade sul-americana difamada em tempos polarizantes, etc., são referências da presença da hostilização utilizada por indivíduos que ainda não aprenderam a respeitar a diferença.

Ademais, tais atitudes se apresentam como uma tentativa de docilização de corpos na disputa de poder entre dominador e dominado (Foucault, 2003), a qual culmina na fragilização uníssona dos colonizados pela datificação hipócrita a favor do sul branqueado europeu masculino heterossexual e que, ainda, precisam conviver com mais esta forma de preconceito perpetuado, agora, digitalmente.

A triste realidade sobre o tema é que precisamos, como diz Silva (2019, 2020), dar nome a esta dor e permanecer na luta visível e invisível. Manter o foco em pedagogias educativas de combate às fobias, alimentadas pelos algoritmos inseridos em IA, é uma alternativa viável para inverter o processo de preconceito e violência, transformando práticas impróprias em ações de cuidado e preservação do indivíduo em suas particularidades enquanto ser humano e detentor de direitos e deveres frente à sociedade global.

E, como menciona Freire (2011), continuar a lutar para não sermos oprimidos, tão pouco opressores. Esta pode ser uma razão de deboche para os que ainda dizem que "não era sério", mas, para muitos, é e causa dor e desconforto reais que devem ser erradicados do mundo real e do mundo digital.

Referências

BARDIN, Laurence. Análise de conteúdo. Lisboa: Edições 70, 1977.

BARROS, Thiane Neves. A influência das imagens de controle na I.A. Generativa. *In:* SILVA, Tarcizio (org.). **Inteligência Artificial Generativa:** discriminação e impactos sociais. Online: Desvelar. Disponível em: desvelar.org. Acesso em: 11 set. 2025.

BENTO, Maria Aparecida Silva. **Pactos narcísicos no racismo:** branquitude e poder nas organizações empresariais e no poder público. Universidade de São Paulo, São Paulo, 2002.

BOLZANI, Isabel. Conar decide arquivar processo contra propaganda que recriou Elis Regina com inteligência artificial. In: **G1 Economia:** Mídia e Marketing. Disponível em: https://g1.globo.com/economia/midia-e-marketing/noticia/2023/08/23/conar-decide-arquivar-processo-contra-propaganda-que-recriou-elis-regina-com-inteligencia-artificial.ghtm. Acesso em: 19 dez. 2025.

BOURDIEU, Pierre. O Poder Simbólico. Tradução: Fernando Tomaz (português de Portugal). 2.ed. Rio de Janeiro: Editora Bertrand Brasil, 1998.

COULDRY, N.; MEJIAS, U. A. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject. *In:* **Television & New Media**, v. 20, n. 4, p. 336–349, 2 set. 2019. Disponível em: https://journals.sagepub.com/doi/10.1177/1527476418796632. Acesso em: 11 set. 2025.

DWIVEDI, Yogesh K. et al. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, v. 71, n. 0268-4012), p. 102642, 2023. DOI: https://doi.org/10.1016/j.ijinfomgt.2023.102642.

FOUCAULT, Michel. Microfísica do poder. 18 ed. Rio de Janeiro: Graal, 2003.

FREIRE, Paulo. Pedagogia do oprimido. 50. ed. Rio de Janeiro: Paz e Terra, 2011.

LUGONES, M. Rumo a um feminismo descolonial. **Estudos Feministas**, Florianópolis, v. 22, n. 3, p. 935-952, set.-dez. 2014.

MAFFESOLI, M. No fundo das aparências. Tradução de Bertha Halpern Gurovitz. 2 ed. Petrópolis, RJ: Vozes, 1996.

MAFFESOLI, M. O conhecimento comum: Compêndio de Sociologia Compreensiva. Tradução: Aluizio Ramos Trinta. Editora Brasiliense S.A., 1988.

PAZERO, Leticia. *Deepfake* X IA: Comercial com imagem de Elis Regina abre discussão sobre perigos no futuro. In: **CNN Brasil.** Disponível em: https://www.cnnbrasil.com.br/entretenimento/deepfake-x-ia-comercial-com-imagem-de-elis-regina-abre-discussao-sobre-perigos-no-futuro/ #goog_rewarded. Acesso em: 19 dez. 2025.

QUIJANO, A. Colonialidade do poder, eurocentrismo e América Latina. In: LANDER, E. (Org.). **A colonialidade do saber**: eurocentrismo e ciências sociais. Perspectivas latino-americanas. Buenos Aires: CLACSO, 2005. p. 227-278.

SILVA, T. **Racismo Algorítmico em Plataformas Digitais: microagressões e discriminação em código**. [s.l: s.n.]. Disponível em: https://lavits.org/wp-content/uploads/2019/12/Silva-2019-LAVITSS.pdf. Acesso em: 11 set. 2025.

SILVA, Tarcízio. Visão Computacional e Racismo Algorítmico: Branquitude e Opacidade no Aprendizado de Máquina. In: **Revista ABPN**, v. 12, p. 428-448, 2020.

STELLANTIS. **Fiat promotes unprecedented crossmedia explosion starring Elvis Presley to launch campaign for new Fiat Strada.** Disponível em: https://www.media.stellantis.com/br-en/fiat/press/fiat-promotes-unprecedented-crossmedia-explosion-starring-elvis-presley-to-launch-campaign-for-new-fiat-strada. Acesso em: 19 dez. 2025.

TYNES, Brendesha M. et al. From Racial Microaggressions to Hate Crimes: A Model of Online Racism Based on the Lived Experiences of Adolescents of Color. Microaggression Theory: Influence and Implications, p. 194-212, 2018.

VIEIRA, Carla. Como os vieses entram na I.A. Generativa. In*:* SILVA, Tarcizio (org.). **Inteligência Artificial Generativa:** discriminação e impactos sociais. Online: Desvelar. Disponível em: https://desvelar.org/inteligencia-artificial-generativa-discriminacao-e-impactos-sociais/. Acesso em: 11 set. 2025.

YIN, Robert K. Estudo de caso: Planejamento e Métodos. 5. ed. Porto Alegre: Bookman, 2015.