

Imagens fragilizadas e preconceito algorítmico latente: estudo de caso de imagens geradas por IA¹

Fernanda Sevarolli Creston Faria Kistemann ²

Resumo expandido Painel Temático

Nesse estudo, observamos como o preconceito e as microagressões se comportam em exemplos reais de estudos de caso que demonstram que os códigos foram executados a partir de um imaginário branco, masculino, heterossexual e europeu com o intuito de ridicularizar as minorias e as massas no uso de IA Generativa.

As microagressões nas plataformas digitais é uma realidade e Silva (2019) demonstra algumas pesquisas sobre o assunto e salienta que o próprio nome parece minimizar algo que, ao contrário, deveria ser alardeado. Microagressões são palavras e situações que são constantemente minimizadas, como “foi só uma piada”, “ah, foi brincadeira”, entre outras, mas que na verdade ferem e mostram o preconceito e o racismo latente na sociedade e na rede de uma forma brutal.

Em relação ao preconceito algorítmico, tratamos do caso específicos da IA Generativa nesse estudo, especificando seu uso e conceito de acordo com 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).

Compreendemos que se trata de um código criado para agrupar informações e criar as imagens de acordo com o que recebe e condiciona, de acordo com seus índices iniciais de programação. Portanto, se refletirmos sobre a IA Generativa e os resultados encontrados nos últimos anos, vamos encontrar o preconceito refletido nas imagens. Isso denota a utilização de parâmetros que coadunam o “branqueamento” no desejo dos resultados em detrimento de outras etnias. Segundo Vieira (2024),

¹ Trabalho apresentado no ET 2. Decolonização, algoritimização e racismo do XVIII Simpósio Nacional da ABCiber – Associação Brasileira de Pesquisadores em Ciberultura. Realização Faculdade Cáster 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.

(...) 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 esses sistemas estruturados em relação à sociedade em si, pois cada qual evidencia uma realidade ilimitada dentro uma rede que, além de sua característica real de infinitas conexões e nós, ainda tem uma face oculta: a possibilidade de esconder seus reais autores.

Para esse estudo, utilizamos a pesquisa qualitativa exploratória através da metodologia de análise de conteúdo (Bardin, 1977) com o procedimento metodológico de estudo de caso (Yin, 2015). Os dados foram colhidos do arquivo pessoal da autora em testagens de IA, nas quais foram detectados comportamentos abusivos indicando preconceito algorítmico, principalmente.

A análise dos dados levou em consideração a observação do poder simbólico e da disputa de poder empreendida em ambientes digitais. Tais disputas foram analisadas levando em consideração o fenômeno do colonialismo de dados, observado por Couldry e Mejias (2018) como ferramenta de motivação do preconceito e microagressões em ambientes digitais, bem como comportamentos que corroboram com essa realidade.

Em 2024, houve um *boom* no aperfeiçoamento e uso de aplicativos de IA Generativa e, inclusive, distribuição e acesso gratuito a vários deles. Nesse interim, alguns técnicos exploraram a alternativa de inserir o rosto de pessoas nas mais diversas situações e corpos. Apresentamos o aplicativo “Toonme” como possibilidade de utilização de animações ou, em inglês, *cartoons*. Experimentamos o referido aplicativo, o qual pode ser encontrado em site do mesmo nome ou na *app store* do Google, conforme publicidade do aplicativo na Figura 1.

Figura 1 – Publicidade/Ícone *Toonme*



Fonte: *Toonme* (2025).

Nessa época, estávamos iniciando o doutorado e resolvemos testar as princesas de contos de fadas nesse aplicativo, as quais fazem parte de nossa pesquisa. Enquanto pessoa parda, esperávamos que, ao solicitar que o nosso rosto fosse colocado no corpo de uma princesa, a IA pudesse trabalhar o contexto e nos transformar, mantendo a cor parda. Foi dessa experimentação que colhemos a primeira figura a ser analisada. A seguir, apresentamos um dos resultados fornecidos pela IA.

Figura 2 – Princesa “Branca”



Fonte: Arquivo pessoal da autora (2025).

O resultado da solicitação foi a criação da Princesa de “A Bela e a Fera” branca, e não parda. O rosto foi utilizado, mas a pele foi alterada. Esse “branqueamento” proposital nos incomodou e começamos a observar o comportamento de outras IA em relação ao uso de imagens, cor da pele, gênero, raça, etc.

Em um outra IA, observamos um comportamento gordofóbico e exagero nos traços latinos. Observe a comparação entre a imagem fornecida pelo “Toonme” em 2024 e a imagem fornecida pelo “ChaGpt” em 2025.

Figura 3 – Toonme em 2024



Fonte: Arquivo pessoal da autora (2025).

Figura 4 – ChaGpt em 2025



Fonte: Arquivo pessoal da autora (2025).

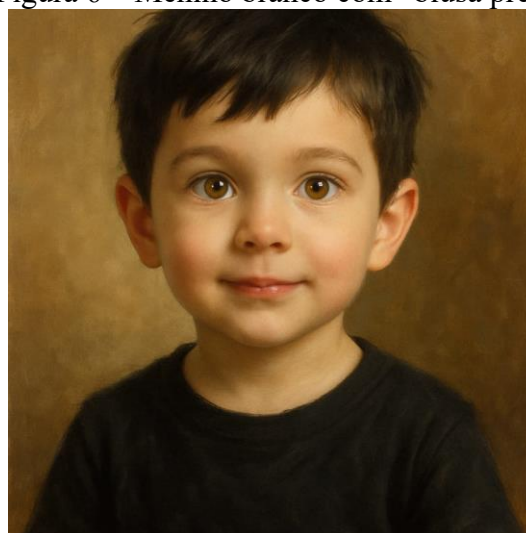
Observando as Figuras 3 e 4, constatamos que as microagressões e o preconceito explícito, demonstrados através das referidas testagens, são os mesmos que ocorrem quando alguém da cor parda ou negra, ou mesmo de outra etnia que não a branca, precisa acessar ou utilizar este tipo de ferramenta digital. O indivíduo vê sua identidade alterada ou grosseiramente acentuada por traços que a IA julga que sejam “melhores”, pois assim foram inseridos ou “aprendidos” por ela.

Outro experimento, também em formato de testagem, foi realizado a partir da provocação de um *post* encontrado na plataforma da rede social do Instagram. Encontramos um *post* no qual uma jovem informava que havia utilizado uma IA, sem mencionar qual, para verificar como seriam seus gatos se fossem crianças, seres humanos. Achamos interessante fazer a mesma tentativa e,

para tanto, utilizamos o ChatGPT e fizemos o mesmo com fotos de gatos: como eles seriam se fossem crianças?

Construímos um *prompt* solicitando que fosse transformado o gato ou a gata em menino ou menina (*Prompt* utilizado: transforme meu gato em um menino ou transforme minha gata em menina). Inserimos na IA 6 (seis) fotos de 6 (seis) gatos de cores diferentes, um de cada vez para respeitar o gênero de cada um. Em meio aos resultados, encontramos dois gatinhos diferenciados que, quando transformados em crianças, não ficaram com suas cores de origem. Observem os resultados dos dois casos a seguir.

Figura 5 – Gato preto transformado em Figura 6 – Menino branco com “blusa preta”



Fonte: Arquivo pessoal da autora (2025).

O ChatGPT se comportou aqui como o aplicativo “Toonme” e tomou a Figura 5 de um gato preto e o transformou em um menino branco, conforme a Figura 6. Podemos dizer que a IA tentou minimizar o problema colocando no menino branco uma blusa preta para que ele não ficasse tão “diferente” do gato preto. Isso até poderia ser considerado uma ironia se não estivéssemos falando de IA.

Contudo não houve ironia alguma ao se colocar no garoto branco uma blusa preta. Os desenvolvedores e reguladores do sistema que desenvolveram o algoritmo do ChatGPT ignoraram a cor do gato, só se atentaram que era um gato macho e que deveriam transformá-lo em um menino.

Figura 7 – Gata mesclada e Figura 8 – Menina com “blusa mesclada”



Fonte: Arquivo pessoal da autora (2025).

Já a Figura 7 traz uma gatinha mesclada que poderia ter sido “transformada” em uma menininha parda ou até mesmo morena. Contudo, o resultado o ChatGPT “transformou” a gatinha em uma menina branca com sardas e, ainda, colocou-a usando uma “blusa mesclada”. Novamente nos deparamos com a ironia dos programadores, desenvolvedores, entre outros, que alimentam e regulam essa IA. Trata-se de uma microagressão juntamente com preconceito algorítmico.

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 IA através da inserção de aprendizagem “maquinica” e códigos de sistema internos automatizados para tanto.

As características que nos surpreenderam pela mudança de identidade visual tanto na princesa quanto nos gatinhos são passadas aos códigos e algoritmos para implantação em plataformas, IA e outros possíveis mecanismos generativos para reprodução em seus sistemas através de seres humanos, reproduzidos a partir do preconceito e da agressão iniciados na mente humana, e não na máquina.

Não há magia ou ironia nesses contextos. Há 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 não são brincadeiras ou piadas, são marcas profundas sociais antigas e humanas.

A simbologia impetrada nos comentários ou na atuação humana através dos algoritmos que motivam microagressões no ambiente digital ainda são dolorosas e causam danos reais nas pessoas. O que parece brincadeira, tem forte cunho simbólico-ideológico destilado na vida das pessoas de pele tingida (Bourdieu, 1998). A colonização dos dados reflete ainda a 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 a marca de gênero, de cor, de classe, de minorias, enfim, marcas humanas de preconceito e agressão que continuam vivas e reificadas nos ambientes virtuais alimentando imaginários (Maffesoli, 1988, 1996) que não deveriam ou não poderiam mais existir.

O branqueamento forçado, a gordofobia em forma de humor, a latinidade americana sendo difamada em tempos polarizantes, etc., são referências hostilizadas e utilizadas por um povo que ainda não aprendeu a respeitar a diferença. É a tentativa de uma docilização de corpos na disputa de poder entre dominador e dominado (Foucault, 2003), que culmina na fragilização uníssona dos que colonizados pela datificação dolosa a favor do sul branqueado europeu masculino heterossexual ainda precisam conviver com mais essa forma de preconceito perpetuado digitalmente.

A triste realidade sobre esse tema é que ainda precisamos, como diz Silva (2019), dar nome a nossa dor e permanecer na luta visível e invisível. Manter o foco em pedagogias educativas de combate às fobias alimentadas pelos algoritmos e IA e, como diz Freire (2011), lutar para não ser oprimido. Essa pode ser uma razão de deboche para os que ainda dizem que “não era sério”, mas para muitos é sério e dói.

Palavras-chave

Preconceito algorítmico; microagressões; IA generativa; imagens.

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.

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.

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.

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.

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.

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: desvelar.org. Acesso em: 11 set. 2025.

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