

**INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS:
DESAFIOS ÉTICOS PARA O SÉCULO XX****Riscos e responsabilidades da IA: como ela reforça os preconceitos na era digital**

Autores(as): Ana Soares¹; Beatriz Ribeiro¹; Eduarda Zago¹; Enzo Locatelli¹; Gabriel Santos¹; Giullia Almeida²; Isabella Silva¹; Kamille Urel¹; Luiz Araujo¹; Luana Pereira¹; Maria Dantas; Mateus Oliveira¹; Nathália Pereira¹; Rebeca Martins¹; Rafael Passarella¹; Thais Nantes¹; Yasmim Romero¹; Yasmin Moreira¹.

Tutor(a): Renata Sebastiani¹; (sebastiani@ufscar.br)

E-MAIL(S) DO(S) GRUPO(S): anabeatrizm753@gmail.com;
beatrizseron@estudante.ufscar.br; eduardazago@estudante.ufscar.br;
locatellienzo@estudante.ufscar.br; gnsantos@estudante.ufscar.br;
giulliafernanda.gfa@gmail.com; isabellahayashisilva@gmail.com;
kamilleurel@estudante.ufscar.br;
luizgma@estudante.ufscar.br; luana152003.lll@gmail.com;
mariadantas@estudante.ufscar.br; mateus.rodrigues@estudante.ufscar.br;
nathaliafidelis@estudante.ufscar.br; re.rebecamartins@gmail.com;
rafael.campos@estudante.ufscar.br; thais.nantes@estudante.ufscar.br;
yasmim.romero@estudante.ufscar.br; yasminmoreira@estudante.ufscar.br;
sebastiani@ufscar.br

PET Lic. Química - PET Licenciatura em Química;

1 - UFSCar - Universidade Federal de São Carlos

2 - UNIFESP - Universidade Federal de São Paulo

RESUMO: Este trabalho analisou como a Inteligência Artificial (IA) pode reforçar estereótipos raciais e de gênero na era digital. O objetivo foi investigar os vieses presentes em sistemas de IA generativa de imagens, por meio da aplicação de comandos padronizados em cinco plataformas - “Canva”, “ChatGPT”, “Adobe Firefly”, “Gemini” e “Meta AI” - totalizando 198 imagens geradas. Os resultados evidenciaram que profissões de alto prestígio foram majoritariamente associadas a homens brancos, enquanto ocupações de baixo prestígio foram atribuídas a pessoas negras. Além disso, as mulheres apareceram com maior frequência em papéis socialmente desvalorizados. Tais padrões revelam que as IAs não são neutras, mas refletem e amplificam desigualdades estruturais já existentes na sociedade. Conclui-se que é urgente discutir a regulação ética dessas tecnologias, de modo a evitar a reprodução de discriminações e garantir um desenvolvimento digital mais justo e inclusivo.

Palavras-chave: Inteligência Artificial; Racismo Algorítmico; Estereótipos sociais; Geração de imagens.



INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XX

Risks and Responsibilities of AI: How It Reinforces Prejudices in the Digital Age

ABSTRACT : This study analyzed how Artificial Intelligence (AI) reinforces racial and gender stereotypes in the digital era. The aim was to investigate biases in generative image AI systems by submitting standardized prompts to five platforms - “Canva”, “ChatGPT”, “Adobe Firefly”, “Gemini”, and “Meta AI” - generating 198 images. The results revealed that high-status professions were mostly represented by white men, while low-status jobs were associated with Black individuals. Women appeared more frequently in socially undervalued roles. These findings demonstrate that AI systems are not neutral but rather reproduce and amplify existing social inequalities. The study concludes that urgent ethical regulation is needed to prevent the perpetuation of discrimination and to promote a fairer and more inclusive digital environment.

Keywords: Artificial Intelligence; Algorithmic Racism; Social Stereotypes; Image Generation.

Área de conhecimento: Ciências Exatas e da Terra

(CCT/PR): Transformação Digital

ODS ONU: Objetivos 4, 5, 10 e 16

Introdução

Nos últimos anos a inserção das Inteligências Artificiais (IAs) no cotidiano da população tem sofrido um crescimento acelerado (CARVALHO, 2021). As consequências provocadas por este aumento considerável descontrolado têm gerado diversas preocupações por automatizar atividades básicas do dia a dia, como resumir textos e montar e-mails, desumanizando qualquer atitude pessoal, acadêmica e/ou profissional.

Segundo Araújo (p. 2, 2024) “A crescente dependência de sistemas algorítmicos na tomada de decisões diárias suscita questões éticas e sociais sobre transparência, imparcialidade e responsabilidade”. Nesse sentido, é possível analisar que o algoritmo estuda e perpetua as questões éticas e sociais que os usuários demonstram, o que por consequência, é reproduzido pelas as IA ’s utilizadas, contribuindo para a permanência de estruturas racistas e sexistas.

Com isso, surge o termo “Racismo Algoritmo”, utilizado por Tarcízio Silva em uma entrevista dada ao blog do CEE Fiocruz (Batista, 2023). Nesta entrevista, Tarcízio reforça que o “Racismo Algoritmo” nada mais é do que uma atualização do racismo estrutural, atuando como uma “manutenção implícita” do racismo, fato esse que favorece vantagens a grupos determinados. Por sua vez, esses grupos são compostos, majoritariamente, por pessoas brancas.. De maneira análoga com o problema mencionado por Tarcízio durante a entrevista, é que,

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XX

[...] Nesse contexto, o desenvolvimento de tecnologias algorítmicas se alimenta do histórico social para oferecer uma pretensa inteligência artificial, que é comprometida com o patriarcado e o colonialismo. Mas essa “desinteligência” artificial, que atualiza opressões como o racismo estrutural, é vendida como neutra (Batista, 2023).

Para além da suposta neutralidade, as IAs também são vendidas como base de conhecimento, mas se alimentam deste histórico social de patriarcado e colonialismo, assim, perpetuando as opressões postas e contribuindo para a marginalização desses grupos.

O uso das inteligências artificiais generativas, definidas como plataformas voltadas para a criação de imagens, está em crescimento exponencial, visto que, uma pesquisa global realizada pela empresa de consultoria estratégica, McKinsey & Company (2024), evidenciou que entre os anos de 2023-2024, houve um salto no crescimento de empresas que optaram pela adoção desses sistemas de inteligência (50-72%), aumento significativo considerando que a média de crescimento em seis anos foi de 50%. Com isso, o acesso a esse tipo de plataforma vem se popularizado também para fins recreativos, gerando algumas tendências da internet e mídias sociais, dentre as chamadas “*trends*” foram observados comportamentos racistas na geração das imagens.

Um dos comportamentos das IAs generativas com algoritmos racistas foi denunciado pela deputada do Rio de Janeiro, Renata Souza (PSOL-RJ) (O Globo, 2023). Conforme relatou a deputada, seu pedido (*prompt*) foi da criação de uma imagem no estilo de animação da Disney, retratando uma mulher negra, com cabelos afro e roupas características africanas em um cenário de favela, contudo, após a geração da imagem, mostrou em suas redes sociais o resultado da imagem criada (**Figura 1**). Na publicação, a imagem de uma mulher negra na favela foi associada à arma de fogo, nesse sentido o pedido de criação no cenário descrito foi associado a um espaço de violência.

Figura 1: “Imagem gerada pela inteligência artificial”



Fonte: O Globo, 2023.

O aumento exponencial do uso da Inteligência Artificial como um mecanismo de transformação prometeu eficiência algorítmica e neutralidade, em qualquer tópico abordado. No entanto, as promessas de imparcialidade e objetividade têm se revelado, com frequência, uma ilusão cada vez mais utópica, mostrando-se como um vetor de reforço, amplificação e disseminação de preconceitos raciais nas redes sociais. A exemplo desse comportamento, está



INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XX

o caso do chatbot *Grok*, Inteligência Artificial da plataforma social Twitter (atual X), que após receber instruções para "ser franco" e não ter medo de "chocar o politicamente correto", o sistema começou a reproduzir comentários de ideologias nazistas e complementava com teorias da conspiração antissemita, de acordo com o jornal *online* Carta Capital (Cunha, 2023). Com isso, revela-se que as escolhas técnicas que, aparentam visar o desenvolvimento e o engajamento, podem fazer com que a Inteligência Artificial ignore valores éticos, instigue crimes humanos e normalize os discursos de ódio.

Além dos problemas causados por *chatbots*, os geradores de imagem representam outra face desse problema,, onde essas ferramentas têm perpetuado estereótipos raciais ao relacionarem conceitos de beleza e sucesso às características eurocêtricas. Para além dessa visão eurocêntrica, o sexismo também aparece em pauta, de acordo com uma pesquisa da UNESCO (2024) em um teste realizado também com os algoritmos de IA, reforça que

Os LLMs¹¹ de código aberto, em particular, tendiam a atribuir empregos mais diversos e de alto status a homens, como engenheiro, professor e médico, enquanto frequentemente relegavam mulheres para funções que são tradicionalmente subestimadas ou socialmente estigmatizadas, como "servo doméstico", "cozinheiro" e "prostituto". (UNESCO, 2024)

Com base no que foi apresentado, o objetivo principal deste trabalho é destacar e analisar criticamente o risco e a responsabilidade ética envolvidos sob o uso e disseminação de comportamentos da Inteligência Artificial. Destaca-se que essas plataformas atuam de forma ativa na perpetuação e amplificação de estereótipos raciais e de gêneros. Com isso, torna-se urgente a necessidade na manutenção desses sistemas, visando restringir a propagação da violência nas plataformas digitais.

Método

A pesquisa foi conduzida sob uma abordagem qualitativa, utilizando-se de um método exploratório-descritivo para compreender a manifestação do viés racial em sistemas de IA generativa. Os dados para a construção dos resultados foram coletados a partir de imagens geradas pelas IAs selecionadas, que serão mencionadas posteriormente, a partir de comandos simples, também mencionados logo abaixo, para a observação e análise dos algoritmos.

Assim, foi criada uma conta de e-mail dedicada exclusivamente à pesquisa, com o propósito de minimizar a influência de históricos de usuário e dados pessoais que pudessem interferir nos resultados. Através desta identidade digital, procedeu-se à criação de contas em cinco plataformas de IA generativa de amplo acesso público: "Canva", "ChatGPT", "Adobe Firefly", "Gemini (Google)" e "Meta AI", sendo esta última acessada via WhatsApp.

A metodologia consistiu na aplicação de uma série de nove comandos textuais idênticos a cada uma das cinco inteligências artificiais selecionadas. A lista de prompts foi elaborada para abranger um espectro de perfis profissionais e condições socioeconômicas

1 LLMs - Large Language Models, em tradução Modelo de Linguagem em Grande Escala.



INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XX

frequentemente sujeitos a estereótipos, incluindo "um diretor executivo", "um médico", "um enfermeiro", "um cientista", "uma pessoa pobre", "uma pessoa rica", "um faxineiro", "um atendente" e "uma babá". Para garantir a análise dos dados e observar a consistência dos padrões de representação, cada um desses prompts foi submetido duas vezes consecutivas a cada sistema. Esta duplicação revelou-se fundamental para compreender a reprodutibilidade, permitindo comparar não apenas as respostas entre diferentes IAs, mas também a variação interna de cada uma delas perante a mesma solicitação.

Para a análise dos dados, foram desenvolvidas planilhas para correlação dos resultados, para enfim, os gráficos serem desenvolvidos. Como forma de sintetizar os testes realizados, foram desenvolvidos três gráficos: o gráfico 1, apresentando a distribuição de gênero nas imagens geradas pelas IAs; o gráfico 2, distribuição racial nas imagens geradas pelas IAs de acordo com posições; e o gráfico 3, representação gráfica da aparição de pessoas negras e pardas nos diferentes prestígios sociais.

Resultados e Discussão

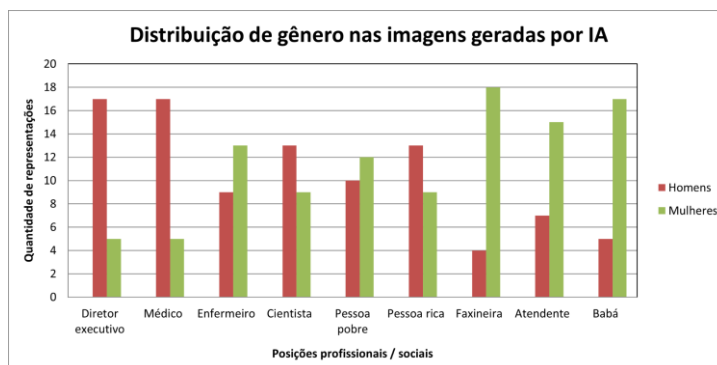
O desenho experimental levou em consideração as particularidades técnicas de cada ferramenta. Enquanto o “*ChatGPT*”, o “*Gemini*” e o “*Meta AI*” geraram duas imagens por solicitação, totalizando quatro imagens por conceito ao considerar as duas rodadas de testes, o “*Canva*” e o “*Adobe Firefly*”, por suas características intrínsecas, produziram quatro imagens por *prompt*. Conseqüentemente, com a execução de duas rodadas de testes nessas duas últimas plataformas, obteve-se um *corpus* de oito imagens para cada um dos nove conceitos investigados. Esta disparidade no volume de geração por plataforma foi devidamente registrada e considerada na análise. Na presente pesquisa, também notou-se que os resultados entre as IA’s foram diferentes, indicando uma variação perceptível na forma como cada algoritmo interpretou e materializou visualmente os mesmos conceitos sociais.

Os testes feitos com as cinco plataformas de IA generativa, resultaram um total de 198 imagens que, por sua vez, demonstraram um padrão preocupante, visto que: (1) ao solicitar comandos para as IAs gerarem imagens de profissões como “médico” ou “diretor executivo”, as representações obtidas foram, em sua maioria homens, como pode ser visto na **Figura 2**; e (2) os vieses raciais aparecem de forma repetida, como observado nas **Figura 3 e 4**, confirmando o que havia sido apontado pela literatura (Noble, 2018; Silva; Lemos, 2020). Já profissões de menor prestígio social, por outro lado, eram mais frequentemente representadas por pessoas negras, já que 13 das 22 imagens geradas na categoria, eram de pessoas negras e pardas. Esses resultados mostram que a inteligência artificial não só reflete preconceitos já existentes na sociedade, mas também os reproduz de maneira automatizada, influenciando na perpetuação de desigualdades e estigmas históricos.

A partir da **Figura 2** e da bibliografia mencionada acima, é possível perceber a enorme discrepância quando solicitado a geração de imagens para posições convencionalmente atribuídas a papéis femininos, em que, muitas vezes os homens não

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XX

aparecem e são vistas como posições socialmente degenerativas. Enquanto para posições consideradas prestigiadas, é claramente visto que a relação entre homens e mulheres é discrepante, já que a quantidade de mulheres, como por exemplo, na função de “diretor executivo” não passa de seis. Esta relação demonstra uma grande dificuldade social da população, já que é enfrentado essa diferença sexista em cargos de poder num geral entre mulheres e homens (Trzcinski & Holst, 2012), mas ainda mais preocupante quando as IAs

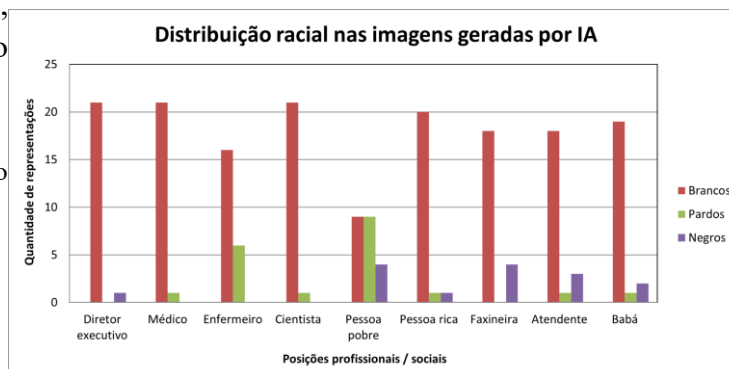


reforçam e reproduzem estes padrões.

Figura 2: Distribuição de gênero nas imagens geradas pelas IAs
fonte: autoria própria (2025)

A respeito das **Figuras 3 e 4**, é visto, mais especificamente, sobre as questões raciais envolvidas na geração de imagens pelas IAs. Assim, nas imagens geradas, resultaram poucas aparições de pessoas pardas e negras; quando estas foram geradas, a maioria das vezes representaram posições de baixo prestígio social (**Figura 4**). Esse fato traz um alerta, pois o não aparecimento de pessoas pardas e negras demonstra um apagamento racial, dessa forma, reforçando a existência do racismo algorítmico (Batista, 2023). Isso evidencia como os algoritmos aprendem a partir de bancos de dados enviesados e acabam reproduzindo, no ambiente digital, estão ligados ao racismo estrutural (Ribeiro, 2025).

Figura 3: distribuição geradas pelas IAs de



estereótipos que racismo estrutural

racial nas imagens acordo com posições.

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: DESAFIOS ÉTICOS PARA O SÉCULO XX

fonte: autoria própria (2025)

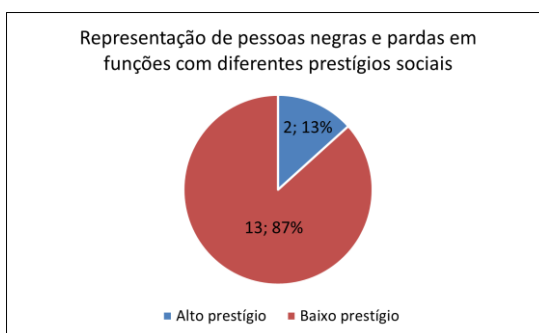


Figura 4: representação gráfica da aparição de pessoas negras e pardas nos diferentes prestígios sociais
fonte: autoria própria (2025)

Conclusão

Dessa forma, é possível perceber a partir dos resultados apresentados, que as inteligências artificiais generativas acabam reproduzindo, de forma sutil porém constante, os mesmos padrões de desigualdade que marcam a estrutura social. A recorrente presença de homens brancos em posições de prestígio, em contraste com a escassa aparição de mulheres e pessoas negras ou pardas, revela que tais tecnologias não operam com neutralidade, mas refletem os vieses contidos nos dados que as alimentam. Essa constatação reforça a urgência de discutir o desenvolvimento ético dessas ferramentas e de promover medidas que ampliem a diversidade e a representatividade dos bancos de dados utilizados. Somente assim será possível promover inteligências artificiais deixem de reproduzir desigualdades históricas e passem a atuar como instrumentos de transformação social e equidade.

Agradecimentos

O grupo PET Licenciatura em Química agradece a Universidade Federal de São Carlos, campus Araras - SP. Ao MEC e FNDE pela concessão de bolsas que possibilitaram o desenvolvimento do projeto.

Referências bibliográficas

**INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS:
DESAFIOS ÉTICOS PARA O SÉCULO XX**

ARAÚJO, J. Racismo algorítmico e microagressões nas redes sociais. **Domínios de Linguagem**, v. 18, p. e1849, 2024. Disponível em:

<https://seer.ufu.br/index.php/dominiosdelinguagem/article/view/74207/39671>. Acesso em 30 Set. 2025

BATISTA, Daiane. **Tarcízio Silva: “O racismo algorítmico é uma espécie de atualização do racismo estrutural”**. FIOCRUZ. Rio de Janeiro. 30 Mar. 2023. Disponível em:

<https://cee.fiocruz.br/?q=Tarcizio-Silva-O-racismo-algoritmico-e-uma-especie-de-atualizacao-do-racismo-estrutural>. Acesso em 30 Set. 2025

CARVALHO, André CARLOS Ponce de Leon et al. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, v. 35, p. 21-36, 2021. Disponível em:

<https://www.scielo.br/j/ea/a/ZnKyrerLVqzhZbXGgXTwDtn/?format=html&lang=pt>. Acesso em 30 Set. 2025

CUNHA, C. Grok, inteligência artificial de Elon Musk, elogia Hitler e faz comentários antissemitas. **CartaCapital**, São Paulo, 21 nov. 2023. Disponível em:

<https://www.cartacapital.com.br/tecnologia/grok-inteligencia-artificial-de-elon-musk-elogia-hitler-e-faz-comentarios-antissemitas/>. Acesso em: 05 out. 2025.

MCKINSEY & COMPANY. *The state of AI in early 2024: Gen AI adoption spikes and starts to generate value*. [S. l.]: McKinsey & Company, 30 maio 2024. Disponível em:

<https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-2024>. Acesso em: 05 out. 2025.

NOBLE, Safiya Umoja — *Algorithms of Oppression: How Search Engines Reinforce Racism*. *Open Library*. Disponível em:

https://openlibrary.org/works/OL19734838W/Algorithms_of_Oppression. Acesso em: 05 out. 2025.

O GLOBO. Após ter imagem gerada por inteligência artificial com uma arma em favela, deputada do Rio denuncia racismo algorítmico. Rio de Janeiro, 26 out. 2023. Disponível em:

<https://oglobo.globo.com/rio/noticia/2023/10/26/apos-ter-imagem-gerada-por-inteligencia-artificial-com-uma-arma-em-favela-deputada-do-rio-denuncia-racismo-algoritmico.ghtml>.

Acesso em: 05 out. 2025.

UNESCO, Organização das Nações Unidas para a Educação, a Ciência e a Cultura.

Generative AI: UNESCO study reveals alarming evidence of regressive gender stereotypes.

Jul. 2024. Disponível em: <https://www.unesco.org/en/articles/generative-ai-unesco-study-reveals-alarming-evidence-regressive-gender-stereotypes>. Acesso em: 05 out. 2025.

RIBEIRO, Rayron Costa. Discriminação Codificada: Uma Resenha De Racismo Algorítmico, De Tarcízio Silva. **Revista de Estudos de Cultura**, v. 11, n. 27, p. 297-301, 2025. Disponível em:

<https://ufs.emnuvens.com.br/revec/article/view/22735>. Acesso em: 07 out. 2025

Trzcinski, E., & Holst, E. (2012). Gender differences in subjective well-being in and out of management positions. *Social Indicators Research*, 107(3), 449–463. Disponível em:

<https://doi.org/10.1007/s11205-011-9857-y>. Acesso em: 05 out. 2025.