

## INTELIGÊNCIA ARTIFICIAL, ÉTICA E RACISMO

Autores(as): <sup>1</sup>RIGAUD, G. L., <sup>2</sup>NETO, C. E. S., <sup>3</sup>LUNA, L. D. M. M., <sup>4</sup>FIDALGO, R. N.;

<sup>1</sup>Grupo PET-Informática, UFPE, Campus SEDE; <sup>2</sup>Grupo PET-Informática, UFPE, Campus SEDE; <sup>3</sup>Grupo PET-Informática, UFPE, Campus SEDE, <sup>4</sup>Tutor(a) do Grupo PET-Informática, UFPE, Campus SEDE

E-mail: [glr2@cin.ufpe.br](mailto:glr2@cin.ufpe.br), [cesn2@cin.ufpe.br](mailto:cesn2@cin.ufpe.br), [ldmml@cin.ufpe.br](mailto:ldmml@cin.ufpe.br), [rdnf@cin.ufpe.br](mailto:rdnf@cin.ufpe.br), [pet@cin.ufpe.br](mailto:pet@cin.ufpe.br);

**RESUMO:** O objetivo do artigo é abordar casos em que a utilização da Inteligência Artificial (IA) falha eticamente, perpetuando preconceitos raciais, discriminação e desigualdade; bem como discorrer sobre os perigos envolvidos nessas falhas e suas consequências.

**Palavras-chave:** viés algorítmico; ética algorítmica; discriminação racial; responsabilidade tecnológica.

### ARTIFICIAL INTELLIGENCE, ETHICS AND RACISM

**ABSTRACT :** The objective of the article is to address cases in which the use of Artificial Intelligence (AI) fails ethically, perpetuating racial prejudices, discrimination, and inequality; as well as to discuss the dangers involved in these failures and their consequences.

**Keywords:** algorithmic bias; algorithmic ethics; racial discrimination; technological accountability.

### Introdução

A presença da Inteligência Artificial (IA) na sociedade contemporânea é inegável, manifestando-se em sistemas de recomendação, assistentes virtuais e modelos de reconhecimento facial. Como qualquer tecnologia amplamente adotada, a IA exerce impacto profundo sobre o cotidiano humano. No entanto, embora frequentemente retratada como neutra e objetiva, ela é construída por processos humanos e, portanto, carrega as marcas de desigualdades e estruturas de poder existentes.

Ao aprender a partir de dados históricos, os algoritmos podem incorporar e amplificar padrões discriminatórios, dando origem aos chamados vieses algorítmicos. Esses vieses comprometem a equidade das decisões automatizadas e produzem consequências éticas e sociais significativas. Casos como o das filter bubbles (Kitchens, Johnson & Gray, 2020) e de falhas em sistemas de reconhecimento facial evidenciam que os algoritmos, quando aplicados a contextos sensíveis — como justiça, crédito ou saúde —, podem perpetuar discriminações e aprofundar desigualdades raciais e estruturais.

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

A suposta neutralidade técnica, portanto, mascara dinâmicas históricas de exclusão, transformando desigualdades sociais em métricas computacionais. Diante disso, torna-se essencial compreender a extensão desses impactos e promover estratégias de mitigação que incluam diversidade nas bases de dados, representatividade nas equipes de desenvolvimento e princípios éticos na governança tecnológica. Esta revisão sistemática da literatura busca, assim, analisar criticamente as manifestações do racismo algorítmico e contribuir para o desenvolvimento de sistemas de IA mais justos, transparentes e socialmente responsáveis.

### Método

Este estudo caracteriza-se como uma revisão sistemática de literatura, cujo objetivo foi investigar como a inteligência artificial apresenta vieses raciais e quais proposições éticas emergem na produção acadêmica. Para tanto, foram adotados os seguintes procedimentos:

- **Fontes de dados:** realizaram-se buscas nas bases Google Scholar e SciELO, selecionadas por sua abrangência em publicações acadêmicas em língua portuguesa e inglesa, bem como por disponibilizarem acesso a textos completos.
- **Recorte temporal:** consideraram-se estudos publicados entre 2016 e 2023, visando mapear transformações recentes nas discussões teóricas e empíricas vinculadas ao racismo algorítmico.
- **Palavras-chave e estratégia de busca:** utilizou-se combinação de termos como “inteligência artificial”, “ética da IA”, “viés algorítmico” e “racismo”, aplicados aos campos de título, resumo e palavras-chave dos documentos. A busca foi restrita a materiais acessíveis em texto completo.
- **Critérios de inclusão:**
  1. Obras que enfoquem explicitamente o racismo ou discriminação racial ligada a sistemas de inteligência artificial;
  2. Trabalhos que apresentem discussão ética da IA — teórica, normativa ou empírica;
  3. Publicações escritas em português ou inglês;
  4. Artigos acadêmicos, dissertações, teses ou relatórios com dados ou reflexão original sobre o tema;
  5. Textos disponíveis integralmente de forma gratuita.
- **Critérios de exclusão:**
  1. Textos que não abordem diretamente o tema central (por exemplo, somente aspectos técnicos de IA sem referência ao viés racial ou ética);
  2. Trabalhos duplicados;
  3. Resenhas genéricas, editoriais ou comunicações curtas que não contenham análise ou dados sobre racismo algorítmico;
  4. Publicações fora do período de 2016-2023;
  5. Textos não disponíveis na íntegra gratuitamente.
- **Processo de seleção:** Triagem inicial feita pela leitura de títulos e resumos para verificar



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

adequação aos critérios de inclusão/exclusão; posteriormente, leitura completa dos textos identificados para confirmação da elegibilidade. Não foi empregado processo com revisores independentes — toda seleção e leitura foram realizadas pelos autores do estudo.

- **Análise de dados e categorização temática:** Os estudos incluídos foram examinados criticamente. Foram extraídas evidências empíricas de viés racial, fundamentos éticos invocados, possíveis mecanismos causais (por exemplo, dados enviesados, ausência de diversidade nas equipes, decisões automatizadas), e as propostas de mitigação sugeridas. Os achados foram organizados em categorias temáticas emergentes — ética, evidência empírica, impactos sociais, soluções — para permitir comparação, síntese crítica e identificação de lacunas na literatura.

### Resultados e Discussão

A análise dos resultados obtidos a partir da revisão sistemática permitiu observar como os sistemas de Inteligência Artificial (IA) reproduzem e amplificam desigualdades raciais já existentes na sociedade. A partir da conceituação de ética, conforme Cortella (2009), compreende-se que ela estabelece os limites da convivência humana e fornece as bases para avaliar os impactos das tecnologias no comportamento social. Dentro desse contexto, a “ética algorítmica” (Rossetti & Angeluci, 2021) emerge como campo essencial para compreender como algoritmos podem incorporar valores e preconceitos humanos. Quando alimentados por dados históricos, esses sistemas herdam vieses e discriminações, o que resulta em decisões enviesadas e potencialmente racistas. Tal discussão torna-se ainda mais relevante diante da crescente autonomia das IAs, reforçando a necessidade de integrar princípios éticos humanos aos processos de desenvolvimento e governança tecnológica.

A partir da Tabela 1 apresentada por Silva (2020), que reúne evidências de vieses raciais em aplicações de visão computacional, foi possível identificar uma série de casos emblemáticos que ilustram a persistência do racismo algorítmico. Entre eles, destaca-se o episódio ocorrido em 2009, quando computadores da Hewlett-Packard (HP) não reconheceram o rosto de um homem negro, embora identificassem facilmente o de uma mulher branca. Outro caso amplamente divulgado foi o do Google Fotos, que em 2015 classificou incorretamente fotos de Jacky Alciné e sua namorada como “gorilas”, evidenciando o viés racial presente no sistema. Episódios semelhantes ocorreram em bancos de imagens, nos quais cabelos crespos foram rotulados como “perucas”, e nas buscas do Google, em que expressões como “tranças bonitas” retornavam majoritariamente imagens de mulheres brancas, enquanto “tranças feias” associavam-se a mulheres negras. Essas ocorrências demonstram que a neutralidade algorítmica

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

é ilusória e que as categorias utilizadas nas bases de dados refletem hierarquias raciais e culturais preexistentes.

**Tabela 1 – Casos Selecionados de Manifestação de Viés/Racismo Algorítmico em Visão Computacional**

Caso / Reação	Data	Problema/Viés Algorítmico	Causa/problema “técnico”
Reconhecimento de faces de computador da HP não reconhece usuário negro	12/2009	Desumanização; Invisibilidade	<ul style="list-style-type: none"> <li>• Base de dados insuficiente</li> <li>• Ausência de testes</li> </ul>
Google marca pessoas negras como gorilas	07/2015	Representação e associação racista; desumanização	<ul style="list-style-type: none"> <li>• Base de dados insuficiente</li> <li>• Base de dados com associações racistas intencionais</li> <li>• Ausência de testes</li> </ul>
Robôs interagentes não encontram rosto de mulher negra	03/2017	Desumanização; Invisibilidade	<ul style="list-style-type: none"> <li>• Base de dados insuficiente</li> <li>• Ausência de testes</li> </ul>
Faceapp embranquece pele para deixar “mais bonita” a selfie	04/2017	Representação eurocêntrica de beleza; desumanização	<ul style="list-style-type: none"> <li>• Base de dados insuficiente</li> <li>• Ausência de testes</li> </ul>
APIs não reconhecem gênero e idade de mulheres negras	02/2018	Representação eurocêntrica de gênero e idade	<ul style="list-style-type: none"> <li>• Base de dados insuficiente</li> <li>• Ausência de testes</li> </ul>
Kairos retira do ar aplicativo de “diversidade”	06/2018	Tipologia racial essencialista	<ul style="list-style-type: none"> <li>• Tipologia centrada nos EUA</li> </ul>
APIs de análise de expressões faciais associam emoções negativas a negros	01/2019	Percepção eurocêntrica; estereotipização	<ul style="list-style-type: none"> <li>• Base de dados insuficiente</li> <li>• Ausência de testes</li> </ul>



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

Fonte: Silva (2020).

O viés racial também se manifestou em redes sociais e sistemas de recomendação. Em 2020, um experimento realizado por usuários do Twitter revelou que o algoritmo da plataforma priorizava imagens de pessoas brancas na pré-visualização de tweets, mesmo quando havia mais fotos de pessoas negras nas publicações (Araujo, 2020). Situação semelhante foi observada no Instagram, onde publicações com fotos de mulheres brancas apresentaram maior engajamento (Ferreira, 2020), demonstrando como a IA influencia a visibilidade e a valorização de corpos brancos. Além disso, estudos indicam que o problema se estende a tecnologias críticas, como os carros autônomos, que apresentaram cerca de 5% menos precisão na identificação de pessoas negras, levantando sérias preocupações sobre segurança (Silva, 2020).

Esses casos evidenciam a influência humana intrínseca na construção de sistemas de IA. Como discutido por WebShare (2021), a IA aprende com grandes volumes de dados e refina seus parâmetros com base nos exemplos fornecidos, refletindo inevitavelmente as percepções e limitações de quem os produz. Isso demonstra que o ideal de neutralidade é falso, pois, conforme apontado por Silva (2020), “a falsa ideia de neutralidade é marcada socialmente pela percepção dos desenvolvedores”. O autor retoma Bento (2002, p. 91) ao afirmar que “a invisibilidade do negro é elemento importante da identidade do branco”, reforçando que a branquitude presente nas equipes de desenvolvimento é determinante na reprodução desses vieses. Dados da U.S. Equal Employment Opportunity Commission (2022) mostram que, embora pessoas negras representem aproximadamente 13% da população dos Estados Unidos, elas ocupam menos de 1% dos cargos executivos nas empresas do Vale do Silício. Essa sub-representação contribui para a manutenção de uma visão hegemônica e homogênea no design das tecnologias, reforçando o racismo estrutural.

O impacto dessas falhas éticas torna-se ainda mais grave quando se considera a aplicação de IAs em setores sensíveis. Na justiça criminal, o caso de Brisha Borden e Vernon Prater (Angwin et al., 2016) ilustra como algoritmos de predição de reincidência penal classificaram incorretamente uma mulher negra como de “alto risco” e um homem branco como de “baixo risco”, revelando vieses sistemáticos que influenciam decisões judiciais. Dois anos depois, as previsões mostraram-se erradas, expondo a falha do sistema. Situação análoga foi relatada por Cara (2019) no campo da saúde, em que um algoritmo desenvolvido pela empresa Optum, nos Estados Unidos, subestimou as necessidades médicas de pacientes negros. Ao usar o custo financeiro como métrica de risco, o sistema desconsiderou o menor acesso de pessoas



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

negras aos serviços de saúde, levando à exclusão de 48.772 diagnósticos crônicos adicionais em comparação com pacientes brancos. Esse exemplo mostra como a ausência de sensibilidade social e racial nos critérios algorítmicos pode comprometer diretamente a equidade no atendimento médico.

Tais achados revelam que, embora a IA seja amplamente promovida como ferramenta de eficiência e precisão, ela pode reforçar desigualdades históricas e perpetuar injustiças quando não submetida a avaliação ética rigorosa. Conforme Silva (2020) discute com base em Epstein (2018), há uma tendência preocupante na comunidade científica: a priorização do desenvolvimento de novos modelos em detrimento da análise crítica de modelos existentes. Isso cria um cenário em que a inovação avança mais rapidamente do que a capacidade de identificar e corrigir falhas estruturais. Em consequência, os riscos sociais da IA racista transcendem a esfera técnica e passam a afetar direitos fundamentais, como segurança, liberdade, oportunidades e saúde.

Esses resultados evidenciam que a Inteligência Artificial reflete as estruturas sociais nas quais é desenvolvida, reproduzindo desigualdades quando não submetida a uma análise ética cuidadosa. A falta de diversidade nos dados e nas equipes de desenvolvimento contribui diretamente para o fortalecimento de vieses raciais, comprometendo a equidade e a confiabilidade desses sistemas. Dessa forma, as discussões apresentadas ressaltam a necessidade de aprofundar o debate sobre as implicações éticas e sociais do uso da IA, destacando que a sua aplicação exige constante reflexão crítica e responsabilidade no enfrentamento das desigualdades reproduzidas tecnologicamente.

### Conclusões

A análise realizada evidencia que os sistemas de Inteligência Artificial, quando desenvolvidos sem uma perspectiva ética e socialmente crítica, podem reforçar o racismo estrutural presente nas relações humanas. Os casos estudados demonstram que o viés algorítmico não é apenas um problema técnico, mas uma consequência direta da reprodução de desigualdades históricas nos dados e nas estruturas de poder das empresas de tecnologia. Assim, torna-se indispensável promover a diversidade nas equipes de desenvolvimento, ampliar a representatividade nos conjuntos de dados e adotar práticas de transparência e auditoria algorítmica. O enfrentamento dessas questões é essencial para que a IA cumpra um papel social responsável e contribua para a construção de tecnologias verdadeiramente justas e inclusivas.

## Agradecimentos

Os autores agradecem ao Ministério da Educação (MEC), por meio do Fundo Nacional de Desenvolvimento da Educação (FNDE), pelo apoio institucional e pelos investimentos destinados ao desenvolvimento das atividades do grupo. Agradecem também aos discentes Bruno Lima e Roseane Gama pela colaboração e empenho na condução deste estudo, bem como aos docentes do Centro de Informática da Universidade Federal de Pernambuco pelo suporte acadêmico e orientações que contribuíram significativamente para a elaboração deste trabalho. Por fim, expressam especial reconhecimento ao petiano Victor Luiz de Paula Lima, pela constante disponibilidade, dedicação nas tratativas de auxílio financeiro e eficaz mediação com o tutor do nosso Grupo.

## Referências

ANGWIN, J. et al. **Machine bias – There’s software used across the country to predict future criminals. And it’s biased against blacks.** 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 15 set. 2023.

ARAUJO, L. **Algoritmo racista? Twitter se defende das acusações.** 2020. Disponível em: <https://propmark.com.br/twitter-se-posiciona-sobre-algoritmo-escolher-rostos-de-pessoas-brancas/>. Acesso em: 17 set. 2023.

CALDAS, A. E. **Racismo nas decisões judiciais: a Inteligência Artificial como uma forma de mitigação.** 2021. Disponível em: <https://repositorio.uniceub.br/jspui/bitstream/prefix/15462/1/AMANDA%20ESPINOLA%20CALDAS%20RA%2021702351.pdf>. Acesso em: 10 set. 2023.

CARA, E. **Como um algoritmo médico nos EUA era preconceituoso com pacientes negros.** 2019. Disponível em: <https://gizmodo.uol.com.br/algoritmo-medico-eua-preconceito-negros/>. Acesso em: 18 set. 2023.

CORTELLA, M. S. **Qual é a tua obra? Inquietações propositivas sobre gestão, liderança e ética.** Petrópolis: Vozes, 2009.

DRUMOND, G. et al. **Racismo algorítmico: o preconceito na programação.** 2020. Disponível em: <https://blogfca.pucminas.br/colab/racismo-algoritmico-o-preconceito-na-programacao/>. Acesso em: 17 set. 2023.

FERREIRA, K. L. **Racismo algorítmico não é apenas sobre engajamento nas redes sociais.** 2020. Disponível em: <https://mundonegro.inf.br/racismo-algoritmico-nao-e-apanas-sobre-engajamento-nas-redes-sociais/>. Acesso em: 17 set. 2023.

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

JEFFERSON, R. **Racial disparity in Silicon Valley: where are the Black executives?** 2020. Disponível em: <https://www.the-future-of-commerce.com/2023/01/31/racial-disparity-in-silicon-valley/>. Acesso em: 18 set. 2023.

NOVAES, L. C. A. **A utilização de softwares de inteligência artificial como ferramenta de perpetuação do racismo institucional.** 2022. p. 119-128. Disponível em: [https://www.cidhcoimbra.com/\\_files/ugd/8f3de9\\_232734456abd43bdb2afeeacd7e632.pdf#page=119](https://www.cidhcoimbra.com/_files/ugd/8f3de9_232734456abd43bdb2afeeacd7e632.pdf#page=119). Acesso em: 10 set. 2023.

RODRIGUES, J. C.; CHAI, C. G. **Inteligência artificial e racismo algorítmico: análise da neutralidade dos algoritmos frente aos episódios de violação de direitos nos meios digitais.** 2023. Disponível em: [https://juslaboris.tst.jus.br/bitstream/handle/20.500.12178/215797/2023\\_rodrigues\\_julia\\_inteligencia\\_artificial.pdf?sequence=1&isAllowed=y](https://juslaboris.tst.jus.br/bitstream/handle/20.500.12178/215797/2023_rodrigues_julia_inteligencia_artificial.pdf?sequence=1&isAllowed=y). Acesso em: 13 set. 2023.

ROSSETTI, R.; ANGELUCI, A. **Ética algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação.** 2021. Disponível em: <https://www.scielo.br/j/gal/a/R9F45HyqFZMpQp9BGTfZnyr#>. Acesso em: 10 set. 2023.

SILVA, T. **Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina.** 2020. Disponível em: <https://abpntrevista.org.br/site/article/view/744/774>. Acesso em: 10 set. 2023.

SILVA, T. **Racismo algorítmico: inteligência artificial e discriminação nas redes digitais.** 2022. Disponível em: <https://books.google.com.br/books?id=BCZcEAAAQBAJ&lpg=PT3&dq=inteligencia%20artificial%20e%20racismo&lr&hl=pt-BR&pg=PP1#v=onepage&q&f=false>. Acesso em: 10 set. 2023.