

INTELIGÊNCIA ARTIFICIAL GENERATIVA NA PRODUÇÃO CIENTÍFICA: AUTORIA, ORIGINALIDADE E VIÉS EM DEBATE — UMA REVISÃO INTEGRATIVA

Generative artificial intelligence in scientific production: authorship, originality, and bias in debate — a literature review

Helen Lavigne Almeida Martins¹, Felipe de Melo Souza²

1. Acadêmica do curso de Medicina da Faculdade de Ciências Médicas de Garanhuns; 2.

Professor orientador do curso de Medicina da Faculdade de Ciências Médicas de Garanhuns.

E-mail: lavignehelen2@gmail.com

Eixo Temático: 02

RESUMO

Introdução: o uso de inteligência artificial (IA) generativa tem se expandido rapidamente nas etapas de busca, síntese, redação e avaliação da produção científica, levantando questões sobre autoria, originalidade e sobre se essa produção pode se prejudicar em razão do viés algorítmico. **Objetivo:** analisar, por meio de revisão de literatura, como o uso de IA generativa reconfigura a autoria e a originalidade na produção científica, aprofundando especificamente se e como o viés associado a esse uso pode prejudicar a própria produção científica, e como o pesquisador deve se posicionar diante desse cenário. **Método:** revisão integrativa da literatura conduzida conforme as etapas de Whitemore e Knafl e de Souza, Silva e Carvalho, com buscas no Google Acadêmico, em bases como Scopus e Web of Science e em periódicos de alto fator de impacto, resultando em corpus de 27 estudos primários, com triagem conduzida por uma única revisora. **Resultados:** verificou-se consenso editorial de que a IA não pode figurar como autora; identificaram-se riscos de fabricação de citações — em um estudo, até 55% das geradas eram inexistentes — e um viés que propomos organizar em quatro camadas: herdado dos dados de treinamento, intrínseco ao modelo, amplificado pelo uso individual (concentração de citações em autores já consolidados) e sistêmico, este último já mensurado diretamente em 41,3 milhões de artigos científicos, nos quais a adoção de IA elevou a produtividade individual mas contraiu em 4,63% o total de tópicos investigados pela ciência. **Conclusão:** a resposta a se a produção científica pode se prejudicar pelo viés é afirmativa e, hoje, diretamente evidenciada, não apenas extrapolada, no nível sistêmico da diversidade temática; o pesquisador deve reter o julgamento crítico e auditar ativamente o viés em todas as suas camadas.

Palavras-chave: Inteligência Artificial. Autoria. Publicações Científicas e Técnicas. Viés. Literatura de Revisão como Assunto.

ABSTRACT

Introduction: the use of generative artificial intelligence (AI) has rapidly expanded across the searching, synthesizing, writing, and evaluation stages of scientific production, raising questions about authorship, originality, and whether that production can itself be harmed by algorithmic bias. **Objective:** to analyze, through a literature review, how the use of generative AI reshapes authorship and originality in scientific production, examining whether and how bias associated with this use can harm scientific production itself, and how the researcher should position themselves. **Method:** an integrative literature review conducted according to the steps

of Whittemore and Knafl and of Souza, Silva and Carvalho, with searches on Google Scholar, on databases such as Scopus and Web of Science, and in high-impact-factor journals, resulting in a corpus of 27 primary studies, with screening conducted by a single reviewer. **Results:** an editorial consensus was found that AI cannot be listed as an author; risks of citation fabrication were identified — in one study, up to 55% of generated citations were nonexistent — together with bias that we propose organizing into four layers: inherited from training data, intrinsic to the model, amplified by individual use (concentration of citations on already established authors), and systemic — this last layer now directly measured across 41.3 million scientific papers, in which AI adoption raised individual productivity while contracting the total range of topics studied by science by 4.63%. **Conclusion:** the answer to whether scientific production can be harmed by bias is affirmative and, today, directly evidenced rather than merely extrapolated at the systemic, topic-diversity level; the researcher must retain critical judgment and actively audit bias across all of its layers.

Keywords: *Artificial Intelligence. Authorship. Scientific and Technical Publications. Bias. Review Literature as Topic.*

INTRODUÇÃO

Desde o lançamento público de LLMs conversacionais, no final de 2022, a inteligência artificial (IA) generativa passou a integrar rapidamente as etapas de busca, síntese, redação e avaliação da produção científica, acompanhada por crescimento acelerado de políticas editoriais sobre esse uso.¹ Isso levanta uma questão epistemológica central: até que ponto um pesquisador que utiliza IA generativa permanece autor pleno do texto, no sentido de deter originalidade e de poder ser responsabilizado por cada afirmação nele contida? A relevância é reforçada por riscos já documentados, como fabricação de citações inexistentes² e reprodução de vieses discriminatórios em sistemas de IA aplicados à ciência.³

Diante da falta de consenso amplo sobre os limites entre apoio legítimo e substituição indevida do julgamento autoral, este trabalho tem como objetivo analisar, por meio de revisão integrativa da literatura, como o uso de IA generativa reconfigura a autoria e a originalidade na produção científica, aprofundando especificamente se e como o viés associado a esse uso pode prejudicar a própria produção científica, e de que forma o pesquisador pode se posicionar de maneira responsável diante desse cenário.

MÉTODO

Trata-se de revisão integrativa da literatura, conduzida conforme as etapas propostas por Whittemore e Knafl⁴ e por Souza, Silva e Carvalho:⁵ identificação da pergunta de pesquisa, busca na literatura, avaliação e categorização dos estudos, análise e interpretação dos dados e apresentação da síntese, com relato do processo de busca inspirado no PRISMA 2020⁶ e orientado por Toronto e Remington.⁷ Preferiu-se esse desenho ao de revisão de escopo por

permitir uma síntese crítica original de fontes heterogêneas, e não apenas mapear a produção sobre o tema.

A pergunta que orientou a busca foi: "De que forma o uso de IA generativa impacta a autoria, a originalidade e a confiabilidade da produção científica, incluindo os riscos de viés, e como o pesquisador deve se posicionar?". As buscas foram realizadas no Google Acadêmico e em bases como Scopus e Web of Science, com descritores booleanos ("artificial intelligence" AND authorship; "large language models" AND bias; "generative AI" AND "research integrity"), complementadas por buscas direcionadas em periódicos de alto fator de impacto (Nature, Science, PNAS) sobre viés e homogeneização por IA, no período de 2021 a 2026.

Foram incluídos artigos revisados por pares, preprints relevantes e posicionamentos institucionais sobre autoria, originalidade, automação de revisões, alucinações, viés ou detecção de texto por IA; a inclusão de preprints seguiu critérios reconhecidos para literatura cinzenta em sínteses de evidência.⁸ A triagem foi conduzida por uma única revisora, limitação discutida adiante. Ao final, 27 estudos primários, majoritariamente de 2023-2026, compuseram o corpus (Quadro 1). Por se tratar de pesquisa bibliográfica, sem dados de seres humanos, dispensa-se aprovação de Comitê de Ética em Pesquisa (CEP/CAAE).

RESULTADOS

Quanto à autoria, periódicos e comitês de ética convergem em vedar a atribuição de autoria à IA, por esta não poder assumir responsabilização por erros ou má conduta, ainda que seu uso como apoio seja amplamente aceito mediante divulgação.^{9,10,11,1} Levantamento com 300 periódicos mostrou que 98,9% dos que possuíam política definida vedavam essa possibilidade, embora 96,6% permitissem uso assistido do tipo ChatGPT.¹

Quanto à originalidade, verificou-se coevolução entre pesquisador, sistema científico e IA,¹² com ganhos em coerência e argumentação, mas risco real de dependência excessiva e perda da voz autoral.¹³ Quanto à automação da busca e síntese, LLMs alcançaram 83% de precisão na extração de dados em revisões sistemáticas,¹⁴ com ferramentas já catalogadas para triagem;¹⁵ todos os estudos alertam para acúmulo de erros em cadeia. Quanto às alucinações, um estudo quantificou fabricação de citações em 55% (GPT-3.5) e 18% (GPT-4).²

Quanto ao viés — e à pergunta central de se a produção científica pode se prejudicar em razão dele —, propomos organizar os achados em quatro camadas articuladas, sistematização própria desta revisão e não uma classificação preexistente em nenhuma fonte isolada. Na camada dos dados, o corpus de treinamento carrega desigualdades geográficas, linguísticas e demográficas da literatura de origem, reproduzidas na síntese^{3,16,17} — limitação estrutural do

próprio paradigma de treinamento em larga escala, e não apenas um risco técnico contornável. Na camada do modelo, há viés de generalização, que leva a IA a exagerar achados científicos,¹⁸ somado a sycophancy — validar em vez de corrigir criticamente.¹⁹ Na camada do uso individual, há amplificação do viés de citação (efeito Mateus), concentrando atenção em autores já consolidados.²⁰

Na camada sistêmica, a evidência deixou de ser apenas extrapolada: em análise de 41,3 milhões de artigos, cientistas que adotam IA tornam-se individualmente mais produtivos (3,02×) e mais citados (4,84×), mas o efeito coletivo é uma contração de 4,63% no total de tópicos investigados pela ciência²¹ — risco que Messeri e Crockett descrevem como "monoculturas de método".²² Em resumos biomédicos, ao menos 13,5% dos de 2024 já mostram vocabulário influenciado por LLM,²³ e a diversidade linguística da escrita científica vem encolhendo desde a popularização desses modelos.²⁴ Mecanismos análogos reforçam o quadro: monocultura algorítmica em sistemas de decisão,²⁵ redução de 10,7% na diversidade de conteúdo entre usuários,²⁶ disparidades de prestígio Norte-Sul mantidas apesar do nivelamento de volume,²⁷ e discriminação linguística deslocada para discriminação por origem percebida.²⁸ Essas evidências sustentam resposta afirmativa: a produção científica pode, sim, se prejudicar pelo viés, sobretudo no nível sistêmico — não apenas no de um artigo isolado.

Quanto à detecção, ferramentas de identificação de texto por IA apresentaram acurácia inferior a 80% em quase todos os casos,²⁹ e a ofuscação reduziu a acurácia de detectores avançados para cerca de 5%.³⁰ Quanto ao futuro da revisão por pares, propõe-se IA na triagem inicial como resposta ao crescimento de 47% nas submissões entre 2016 e 2022, mantendo-se a decisão final humana;³¹ cautela reforçada por evidência de que revisões assistidas por IA no ICLR pontuaram mais alto, elevando em 31% as chances de aceitação de artigos limítrofes.³²

O Quadro 1 sintetiza, estudo a estudo, os 27 achados primários incluídos nesta revisão.

Quadro 1 — Síntese comparativa dos estudos incluídos na revisão.

#	Autor(es)/Ano	Principal achado
1	Lund; Naheem (2023)	98,9% dos periódicos com política vedam IA como autora; 96,6% permitem uso assistido.
2	Walters; Wilder (2023)	55% (GPT-3.5) e 18% (GPT-4) das citações geradas eram fabricadas.
3	Afreen et al. (2025)	Viés societal, cognitivo e algorítmico em IA generativa; mitigação técnica e não técnica.
9	Nature (2023, editorial)	Nenhum LLM pode ser autor creditado; exige divulgação do uso.
10	COPE (2024)	IA não pode ser autora por não poder responder por má conduta ou conflitos de interesse.
11	ICMJE (2024)	Chatbots não podem ser autores; uso deve ser divulgado e revisado integralmente pelos autores.
12	Watson et al. (2025)	Autoria como acoplamento autopoietico entre pesquisador, sistema científico e IA.
13	Sanz-Tejeda et al. (2026)	Ganhos em coerência textual; risco de dependência e perda de voz autoral.
14	Scherbakov et al. (2025)	83% de precisão na extração de dados em revisões; fraqueza em dados numéricos.
15	Ge et al. (2024)	Catálogo de ferramentas de IA para revisões; alerta para erro acumulado em cadeia.
16	Suenghataiphorn et al. (2025)	Viés de dados e de modelo em LLMs clínicos; dano alocativo/representacional.
17	Bender et al. (2021)	LLMs replicam e amplificam vieses sociais dos dados; corpus de treinamento não é auditável.
18	Peters; Chin-Yee (2025)	LLMs exageram/generalizam achados científicos além do que os dados sustentam.
19	Sharma et al. (2023)	Sycophancy: LLMs tendem a validar o usuário em vez de corrigir criticamente.
20	Algaba et al. (2024)	LLMs amplificam padrões humanos de citação (efeito Mateus).

#	Autor(es)/Ano	Principal achado
21	Hao; Xu; Li; Evans (2026)	IA eleva produtividade individual (3,02×), mas contrai 4,63% o foco coletivo da ciência.
22	Messeri; Crockett (2024)	Risco de "monoculturas de método" na ciência assistida por IA.
23	Kobak et al. (2025)	≥13,5% dos resumos biomédicos de 2024 mostram vocabulário influenciado por LLM.
24	Sourati et al. (2026)	IA reduz em até 50% a variância de complexidade textual em resumos científicos.
25	Kleinberg; Raghavan (2021)	Uso do mesmo algoritmo por múltiplos agentes pode piorar o resultado coletivo (monocultura).
26	Doshi; Hauser (2024)	IA eleva qualidade individual, mas aumenta 10,7% a similaridade entre textos.
27	Ni; Qu; Wang (2026)	IA reduz distância Norte-Sul em volume, mas mantém disparidades de prestígio.
28	Lepp; Smith (2025)	Discriminação linguística desloca-se para discriminação por origem percebida.
29	Weber-Wulff et al. (2023)	Detectores com acurácia < 80%; texto parafraseado escapa em 74% dos casos.
30	Pellegrina; Helmy (2025)	Ofuscação reduz acurácia de detectores de quase 100% para ~5%.
31	Bauchner; Rivara (2024)	Triagem inicial por IA na revisão por pares; decisão final humana.
32	Russo et al. (2025)	Revisões assistidas por IA no ICLR: +31% nas chances de aceitação de artigos limítrofes.

Fonte: elaborado pelos autores (2026).

CONCLUSÃO

Como limitações: triagem por uma única revisora, sem dupla checagem; dependência de preprints em parte da evidência sobre viés; e caráter recente do objeto, que ainda limita réplicas dos achados mais novos da camada sistêmica.

A leitura integrada dos achados delinea como o pesquisador deve se posicionar diante do uso de IA. A divulgação do uso é condição mínima de transparência; toda citação sugerida por IA deve ser verificada de forma independente, dado o risco de fabricação;² decisões de relevância devem permanecer sob julgamento humano, pois o erro acumulado em processos automatizados é risco já demonstrado;^{14,15} e a auditoria de viés precisa verificar as quatro camadas propostas — dados, modelo, uso individual e efeitos sistêmicos.^{3,16-28} Como os detectores de texto por IA carecem de confiabilidade,^{29,30} essa integridade depende, hoje, mais da honestidade do pesquisador do que de verificação técnica automática.

Este trabalho propôs uma sistematização em quatro camadas para os riscos de viés na produção científica assistida por IA, reunindo evidência direta e de alto impacto de que essa adoção já contrai a diversidade temática da ciência. A produção científica caminha para uma reorganização híbrida do trabalho, em que a IA assume funções de apoio, e o pesquisador retém a formulação do problema, o julgamento crítico e a responsabilização final. Sugere-se que estudos futuros repliquem, com dupla checagem independente, essa mensuração.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Lund BD, Naheem KT. Can ChatGPT be an author? A study of artificial intelligence authorship policies in top academic journals. *Learned Publishing*. 2023.
2. Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci Rep*. 2023;13:14045.
3. Afreen J, Mohaghegh M, Doborjeh M. Systematic literature review on bias mitigation in generative AI. *AI Ethics*. 2025;5(5):4789-841.
4. Whittemore R, Knafl K. The integrative review: updated methodology. *J Adv Nurs*. 2005;52(5):546-53.
5. Souza MT, Silva MD, Carvalho R. Revisão integrativa: o que é e como fazer. *Einstein (São Paulo)*. 2010;8(1):102-6.
6. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
7. Toronto CE, Remington R, editors. *A step-by-step guide to conducting an integrative review*. Cham: Springer; 2020.

8. Hong QN, Fàbregues S, Bartlett G, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Educ Inf.* 2018;34(4):285-91.
9. Nature. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use [editorial]. *Nature.* 2023;613(7945):612.
10. Committee on Publication Ethics (COPE). Authorship and AI tools. London: COPE; 2024. Disponible em: publicationethics.org.
11. International Committee of Medical Journal Editors (ICMJE). Recommendations: use of AI by authors. [S. l.]: ICMJE; 2024. Disponible em: [icmje.org](https://www.icmje.org).
12. Watson S, Brezovec E, Romic J. The role of generative AI in academic and scientific authorship: an autopoietic perspective. *AI Soc.* 2025;40(5):3225-35.
13. Sanz-Tejeda A, Domínguez-Oller JC, Baldaquí-Escandell JM, et al. The impact of generative AI on academic reading and writing: a synthesis of recent evidence (2023-2025). *Front Educ.* 2026;10:1711718.
14. Scherbakov D, Hubig N, Jansari V, et al. The emergence of large language models as tools in literature reviews: a large language model-assisted systematic review. *J Am Med Inform Assoc.* 2025;32(6):1071-86.
15. Ge L, Agrawal R, Singer M, et al. Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. *Syst Rev.* 2024;13:269.
16. Suenghataiphorn T, Tribuddharat N, Danpanichkul P, et al. Bias in large language models across clinical applications: a systematic review. *arXiv.* 2025;2504.02917.
17. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the dangers of stochastic parrots: can language models be too big? *FACCT '21. ACM;* 2021. p.610-23.
18. Peters U, Chin-Yee B. Generalization bias in large language model summarization of scientific research. *R Soc Open Sci.* 2025;12(4):241776.
19. Sharma M, Tong M, Korbak T, et al. Towards understanding sycophancy in language models. *arXiv.* 2023;2310.13548.
20. Algaba A, Mazijn C, Holst V, et al. Large language models reflect human citation patterns with a heightened citation bias. *arXiv.* 2024;2405.15739.
21. Hao Q, Xu F, Li Y, et al. AI tools expand scientists' impact but contract science's focus. *Nature.* 2026;649(8099):1237-43.
22. Messeri L, Crockett MJ. Artificial intelligence and illusions of understanding in scientific research. *Nature.* 2024;627(8002):49-58.
23. Kobak D, González-Márquez R, Horváth EÁ, et al. Delving into LLM-assisted writing in biomedical publications. *Sci Adv.* 2025;11(27):eadt3813.
24. Sourati Z, Karimi-Malekabadi F, Ozcan M, et al. The shrinking landscape of linguistic diversity in the age of LLMs. *Nat Hum Behav.* 2026.
25. Kleinberg J, Raghavan M. Algorithmic monoculture and social welfare. *Proc Natl Acad Sci U S A.* 2021;118(22):e2018340118.
26. Doshi AR, Hauser OP. Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Sci Adv.* 2024;10(28):eadn5290.
27. Ni R, Qu Y, Wang J. Bridging or widening the gap? AI's role in global research performance. *J Chin Polit Sci.* 2026;1-39.
28. Lepp H, Smith DS. "You cannot sound like GPT": language discrimination in CS publishing. *FACCT '25. ACM;* 2025.
29. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, et al. Testing of detection tools for AI-generated text. *Int J Educ Integr.* 2023;19:26.
30. Pellegrina D, Helmy M. AI for scientific integrity: detecting breaches, errors, and misconduct. *Front Artif Intell.* 2025;8:1644098.
31. Bauchner H, Rivara FP. Use of artificial intelligence and the future of peer review. *Health Aff Sch.* 2024;2(5).
32. Russo G, Horta Ribeiro M, Davidson TR, et al. The AI review lottery. *Proc ACM HCI.* 2025;9(CSCW):3757667.