Science in a Closed Loop: When AI Writes, Cites and Validates Itself

A Ciência em Circuito Fechado: Quando a IA Escreve, Cita e Valida a Si Própria

Helena Donato[1,2]

Atualmente, a inteligência artificial (IA) é frequentemente apresentada como uma ameaça súbita à publicação científica. No entanto, a realidade é mais complexa: a IA não criou as fragilidades deste sistema, mas potenciou-as, tornando-as mais rápidas, mais baratas e mais difíceis de controlar. Práticas como a produção fraudulenta de artigos, a fabricação de dados, o plágio, a manipulação de citações, a compra de autoria e a pressão para publicar já existiam antes da IA.

Os LLMs (*large language models*) conseguem gerar revisões da literatura, redigir manuscritos e criar referências convincentes em segundos.[1,2] Muitas dessas referências são falsas. Outras existem, mas não correspondem ao conteúdo citado. Estudos como os recentemente publicados na Lancet[3] e na Nature[4] mostram precisamente isso: referências aparentemente legítimas que remetiam para artigos inexistentes ou para trabalhos completamente diferentes.

A integridade científica sempre dependeu de um pressuposto simples: aquilo que é citado existe, foi lido e sustenta efetivamente a afirmação apresentada. Hoje, esse pressuposto deixou de poder ser tomado como garantido.

O estudo publicado na Lancet analisou 2,5 milhões de artigos biomédicos e identificou milhares de referências fabricadas, muitas delas plausíveis, corretamente formatadas e atribuídas a investigadores reais.[3]

O artigo *The Growing Problem of Hallucinated Citations*, publicado na *Nature*, analisa o crescimento rápido das referências bibliográficas inventadas ou adulteradas geradas por modelos de inteligência artificial (IA), mostrando que este problema já se tornou uma ameaça concreta à integridade da publicação científica.[4]

As referências falsas parecem credíveis porque combinam elementos autênticos de artigos reais. Naddaf M, *et al*[3] referem-se a estas referências como "Frankenstein citations", construções híbridas em que autores reais, títulos plausíveis, revistas existentes e DOIs parcialmente corretos são misturados para criar uma citação fictícia, mas convincente.

Estes erros são frequentemente fáceis de detetar, mas o problema é que os revisores raramente verificam cada referência individualmente. Mas ainda mais preocupante, é que alguns revisores estão a delegar a avaliação dos manuscritos em ferramentas de IA, apesar das revistas não o permitirem devido ao risco de erros, enviesamentos e falhas de confidencialidade. Ao transferir para sistemas automatizados uma responsabilidade que deveria assentar no julgamento crítico humano, esvazia-se o próprio sentido da revisão por pares.[2,5,6]

A inclusão de referências inventadas ou "alucinadas" por sistemas de IA constitui uma violação grave da integridade científica. Quando estas referências fabricadas são detetadas durante o processo editorial ou de revisão por pares, são consideradas um indicador potencial de fraude científica mais ampla, e o manuscrito é geralmente rejeitado. No entanto, se só forem identificadas após a publicação, podem levar à retratação do artigo, comprometendo a credibilidade do trabalho e deixando um impacto reputacional duradouro associado aos autores.

Mas a questão central já não é apenas a existência de referências fabricadas. O verdadeiro perigo é a possibilidade de entrarmos num ciclo fechado de produção de conhecimento, em que textos gerados artificialmente passam a alimentar modelos futuros sem validação humana substantiva. Estamos, assim, perante um risco novo: um ecossistema científico em que a IA escreve os artigos, a IA ajuda a revê-los, a IA resume-os. E, progressivamente, a IA aprende a partir deles.

Mas onde fica o ser humano neste circuito?

O cenário parece extremo, mas já não é teórico. ■

Ethical Disclosures

Conflicts of Interest: The authors have no conflicts of interest to declare.
Financial Support: This work has not received any contribution grant or scholarship.
Provenance and Peer Review: Commissioned; without external peer review.

Responsabilidades Éticas

Conflitos de Interesse: Os autores declaram a inexistência de conflitos de interesse.
Apoio Financeiro: Este trabalho não recebeu qualquer subsídio, bolsa ou financiamento.
Proveniência e Revisão por Pares: Solicitado; sem revisão externa por pares.

[1]Serviço Documentação e Informação Científica, Hospitais da Universidade de Coimbra, Unidade Local de Saúde de Coimbra, Coimbra, Portugal.
[2]Editora Técnica, SPMI Case Reports

https://doi.org/10.60591/crspmi.598

Corresponding Author/ Autor Correspondente:
Helena Donato - helena.donato@gmail.com
Serviço Documentação e Informação Científica, Hospitais da Universidade de Coimbra, Unidade Local de Saúde de Coimbra;
Praceta Prof. Mota Pinto
3004-561 Coimbra, Portugal

Received / Recebido: 27/05/2026
Accepted / Aceite: 28/05/2026
Published online / Publicado online: 31/07/2026
Published / Publicado: 31/07/2026

REFERÊNCIAS

1. Ibrahim El, Voyer A. Qualitative research with LLM chatbots: Technological reflexivity for interpretative technology. Qualitative Res. 2025;26:133-59. doi: 10.1177/14687941251390794
2. Donato H. Inteligência Artificial na Publicação Científica: Implicações para Autores, Revisores e Editores. Rev Port Med Interna. 2025;32:234-7.doi: 10.24950/rspmi.2846.
3. Topaz M, Roguin N, Gupta P, Zhang Z, Peltonen LM. Fabricated citations: an audit across 2·5 million biomedical papers. Lancet. 2026;407:1779-81. doi: 10.1016/S0140-6736(26)00603-3.
4. Naddaf M, Quill E. Hallucinated citations are polluting the scientific literature. What can be done? Nature. 2026;652:26-9. doi: 10.1038/d41586-026-00969-z.
5. Flanagin A, Kendall-Taylor J, Bibbins-Domingo K. Guidance for Authors, Peer Reviewers, and Editors on Use of AI, Language Models, and Chatbots. JAMA. 2023;330:702-3. doi: 10.1001/jama.2023.12500.
6. Metze K, Morandin-Reis RC, de Ávila Reis MF, da Silva Fago M, Florindo JB. Misinformation, false positives and delegation of tasks - Large Language Models should not be used for the detection of retracted literature - A study of 21 Chatbots. J Clin Anesth. 2025;107:112032. doi: 10.1016/j.jclinane.2025.112032.