

Use of generative artificial intelligence in medical research

Policies must be standardised to ensure accountability and maintain public trust

Nazrul Islam, associate professor, ¹ **Mihaela van der Schaar**, professor ²

¹ Faculty of Medicine, University of Southampton, Southampton, UK

² Cambridge Centre for AI in Medicine, Cambridge, UK

Correspondence to: N Islam Nazrul.Islam@soton.ac.uk

Among the most groundbreaking advancements in artificial intelligence (AI) is the emergence of generative AI. These models, including tools such as ChatGPT and Bard, harness deep learning techniques to interpret and replicate complex patterns found in vast datasets. Their primary function is to generate original text or images, demonstrating an understanding of intricate associations, such as the context and semantics of language. For instance, chatbots such as ChatGPT interpret user inputs, typically sentences or paragraphs, to craft unique, contextually relevant responses. This capability stems from training on diverse and extensive textual and imaging data, enabling the model to predict and generate language and images with high accuracy. ChatGPT had 100 million monthly users within the first two months of its introduction, showing the considerable impact and appeal of generative AI.¹

As expected, the use of generative AI tools in academic research and writing has been increasing.¹ In response, many academic publishers and journals, including The BMJ, have published and/or updated their policies around the responsible use of generative AI in academic research and writing.²⁻⁵ In a linked paper, however, Ganjavi and colleagues (doi:10.1136/bmj-2023-077192) found a worrying lack of guidance on the use of generative AI among the top publishers and science journals.⁶ They analysed 100 of the largest publishers (based on the number of journals published), of which more than three quarters (76%) did not provide any guidance on the use of generative AI. Only 15 of the largest 25 publishers had any guidance on the use of generative AI.

The authors also analysed the 100 highest ranked scientific journals, as determined by the H (Hirsch) index. The H index is a metric for assessing a journal's impact, often considered more stable than the journal impact factor.⁷ Ganjavi and colleagues found that 87% of these top journals have guidelines concerning the use of generative AI. Notably, however, these guidelines varied substantially across leading publishers and journals; this variation extended to both the application of generative AI in research and the protocols for reporting its usage.

The discrepancies underscore the emerging nature of generative AI in academic publishing and the evolving standards surrounding its use.

Standardise policy

Publishers and journals have differing policies on the formatting of manuscripts and reporting of research. However, most top journals agree on some fundamental aspects of research conduct and reporting (eg, authorship criteria, standardised reporting guidelines such as CONSORT, STROBE, PRISMA) as specified by the International Committee of Medical Journal Editors.⁸ This is to ensure that ethical principles are followed in attributing authorship and that research is reported to improve transparency and reproducibility.^{9,10} Given the surge in retraction of published research over the past decade, with more than 10 000 papers retracted in 2023 alone,¹¹ more measures are needed to ensure a shared responsibility in minimising the misconduct of research.

Research misconduct hurts everyone—authors, funders, journals, publishers, and, more importantly, public trust in scientific evidence. Generative AI is likely to facilitate many aspects of research, such as improving the quality of writing or helping with finding and summarising useful related information, but it is also likely to make it challenging to reaffirm that rigorous steps were taken to ensure the highest standards in scientific research.

For instance, one recent study reported that 18-55% of the references in AI generated bibliographies were fabricated, and that 24-43% of the non-fabricated references had substantial citation errors.¹² This issue is particularly pressing with the proliferation in conspiracy theories, misinformation, disinformation, and scepticism towards scientific consensus, such as the antivax movement. It is imperative that the scientific community proactively adopts measures to ensure ethical use and responsible reporting of research, particularly when employing generative AI tools.

Shared accountability framework

To uphold standardised reporting, as well as ethical and responsible use of generative AI, particularly in medical journals, a multifaceted approach is needed. First and foremost, comprehensive guidelines that clearly outline the acceptable use of generative AI in research should be developed and implemented. These guidelines should focus on ensuring accuracy, transparency, and reproducibility of study results. Secondly, medical journals must implement peer review processes tailored to identify and scrutinise AI generated content to safeguard against inaccuracies and ethical lapses. Thirdly, collaboration between clinical

scientists, editorial boards, AI developers, researchers, and ethical boards is needed to foster an understanding of the capabilities and limitations of these tools. Better collaboration would generate AI systems that support the clinical community while aligning fully with the ethical standards of scientific research. Fourthly, educational initiatives should be implemented to raise awareness among researchers about the potential pitfalls and ethical considerations of using generative AI. Fifthly, a robust framework for transparency and accountability must be established, whereby researchers using generative AI are required to disclose the extent of AI involvement in their work. Lastly, ongoing research into the impacts of AI on scientific integrity will be instrumental in continually refining these strategies, ensuring they remain effective and relevant in the rapidly evolving AI landscape.

Through these efforts, the scientific community can harness the benefits of generative AI while maintaining the highest standards of research integrity and ethical responsibility, and ensure that the trust in scientific publications, particularly in the specialty of medicine, remains unshaken.

Competing interests: The BMJ has judged that there are no disqualifying financial ties to commercial companies. The authors declare the following other interests: NI is a research editor at *The BMJ*.

References

- 1 Tang A, Li K-K, Kwok KO, *et al*. The importance of transparency: Declaring the use of generative artificial intelligence (AI) in academic writing. *J Nurs Scholarsh*. 2023;Epub ahead of print. doi: 10.1111/jnu.12938
- 2 Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature*. 2023;613:620–1.
- 3 Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379:313.
- 4 Macdonald H, Abbasi K. Riding the whirlwind: BMJ's policy on artificial intelligence in scientific publishing. *BMJ*. 2023;382:p1923.
- 5 Koller D, Beam A, Manrai A, *et al*. Why We Support and Encourage the Use of Large Language Models in NEJM AI Submissions. *NEJM AI*. 2023;1:AIe2300128.
- 6 Gangavi C, Eppler MB, Pekcan A, *et al*. Publishers' and journals' instructions to authors on use of generative artificial intelligence in academic and scientific publishing: bibliometric analysis. *BMJ* 2024;384:e077192.

- 7 Delgado-López-Cózar E, Cabezas-Clavijo Á. Ranking journals: could Google Scholar Metrics be an alternative to Journal Citation Reports and Scimago Journal Rank? *Learn Publ.* 2013;26:101–14.
- 8 ICMJE. Recommendations | Preparing a Manuscript for Submission to a Medical Journal. <https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html> (accessed 3 January 2024)
- 9 Groves T, Godlee F. Open science and reproducible research. *BMJ.* 2012;344:e4383.
- 10 Islam N, Cole TJ, Ross JS, *et al.* Post-submission changes to prespecified statistical analysis plans. *BMJ.* 2022;378:o2244.
- 11 Van Noorden R. More than 10,000 research papers were retracted in 2023 — a new record. *Nature.* 2023;624:479–81.
- 12 Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci Rep.* 2023;13:14045. doi: 10.1038/s41598-023-41032-5