Beyond detection: The quieter risks of large language models in scientific publishing

Ömer Asan[1]
[1]M.D., Council of Forensic Medicine, Istanbul, Turkiye
https://orcid.org/0000-0002-4340-4553

Dear Editor,

The recent editorial by Uzman Ozbek (1) provides a careful account of the ethical challenges that large language models (LLMs) now pose for scientific publishing, including the limited reliability of detection tools, the disproportionate flagging of non-native English authors as suspected AI users (2), and the impossibility of treating non-human systems as accountable authors (3). I would like to extend this discussion in a direction that deserves more editorial attention: the risks that remain even when authors disclose their AI use openly and avoid overt misconduct such as fabricated data or invented citations (4).

Most current debate focuses on what can, in principle, be caught. Yet a subtler problem may matter more in the long run: the gradual weakening of the fit between evidence, uncertainty, and claim that polished, model-assisted prose can produce without anyone noticing.

Several recent findings make the concern concrete. When LLMs summarise scientific work they overgeneralise systematically, nearly five times more often than human writers (5). They often sound confident when wrong, and readers mistake decisive tone for evidential strength (6). Even when AI-suggested references are real and on-topic, they may only loosely support — or contradict — the claims they accompany; a medical evaluation found 50–90% of LLM responses not fully grounded in the sources they cited (7). Lexical analyses of fifteen million PubMed abstracts already detect stylistic traces of LLM-assisted writing (8), and co-writing studies show that model assistance reduces lexical and content diversity (9). Redundant and low-yield publications have risen alongside LLM adoption (10–12). When models advise on methods they often recommend additional complexity, though current reporting guidance for clinical prediction and machine learning emphasises transparency, applicability, and decision relevance over technical sophistication (13,14). LLMs are now used at both ends of the publishing pipeline, with evidence that automated reviewers favour model-shaped prose (15,16), rewarding the same drift twice.

Considered separately, these effects look minor. Considered together, they describe a slow erosion of calibration in the scientific record. A cautious “may suggest” upgraded to “demonstrates” can travel a long way once a paper is cited, summarised, or translated into clinical recommendations, and unlike fabricated references such drift leaves no clean signature.

The editorial closes with the question of whether human and machine writing can still be distinguished (1). A companion question may be more tractable and more useful: even when disclosure is honest, are editorial and peer-review processes equipped to assess what AI has done to the argument? Asking authors not only whether AI was used but for what purpose, and prompting reviewers to test whether each claim remains proportionate to the design, sample, and uncertainty (3,17),

DOI: 10.5505/kpd.2026.10586

Cite this article as: Asan O. Beyond detection: The quieter risks of large language models in scientific publishing. Turkish J Clin Psych 2026; 29:159-160

The arrival date of article: 28.05.2026, **Acceptance date publication:** 15.06.2026

Turkish J Clinical Psychiatry 2026;29:159-160

This work is licensed under Creative Commons Attribution-NonCommercial NonDerivatives International License

would move the editorial task from a futile contest with detection tools toward the evaluative judgement that remains uniquely human. Such structured prompts would also protect authors who use AI fairly, including non-native English speakers, from being penalised for the wrong reasons (2). In an era of fluent machines, the most important safeguard for the scientific record may not be better detection, but better questions.

Correspondence address: M.D., Omer Asan, Council of Forensic Medicine, Istanbul, Turkiye
omerasan@hotmail.com

REFERENCES

1. Uzman Ozbek S. Scientific publishing in the era of artificial intelligence. Turkish J Clin Psych. 2026;29(1):5–7. doi:10.5505/kpd.2026.94580.

2. Pratama AR. The accuracy-bias trade-offs in AI text detection tools and their impact on fairness in scholarly publication. PeerJ Comput Sci. 2025;11:e2953. doi:10.7717/peerj-cs.2953.

3. Zielinski C. Why artificial intelligence is not an author. Eur Sci Editing. 2025;51:e142904. doi:10.3897/ese.2025.e142904.

4. Elali FR, Rachid LN. AI-generated research paper fabrication and plagiarism in the scientific community. Patterns (N Y). 2023;4(3):100706. doi:10.1016/j.patter.2023.100706.

5. Peters UH, Chin-Yee B. Generalization bias in large language model summarization of scientific research. R Soc Open Sci. 2025;12(4):241776. doi:10.1098/rsos.241776.

6. Zhou K, Hwang JD, Ren X, Sap M. Relying on the unreliable: the impact of language models' reluctance to express uncertainty. In: Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Bangkok: Association for Computational Linguistics; 2024. p. 3623–43. doi:10.18653/v1/2024.acl-long.198.

7. Wu K, Wu E, Wei K, Zhang A, Casasola A, Nguyen T, Riantawan S, Shi P, Ho D, Zou J. An automated framework for assessing how well LLMs cite relevant medical references. Nat Commun. 2025 Apr 16;16(1):3615. doi: 10.1038/s41467-025-58551-6. PMID: 40240349; PMCID: PMC12003634.

8. Kobak D, González-Márquez R, Horvát EÁ, Lause J. Delving into LLM-assisted writing in biomedical publications through excess vocabulary. Sci Adv. 2025;11(27):eadt3813. doi:10.1126/sciadv.adt3813.

9. Padmakumar V, He H. Does writing with language models reduce content diversity? In: Proceedings of the 12th International Conference on Learning Representations (ICLR); 2024 May 7–11; Vienna, Austria. Available from: https://openreview.net/forum?id=Feiz5HtCD0

10. O'Grady C. Low-quality papers surge thanks to public data and AI. Science. 2025;388(6749):807–8. doi:10.1126/science.adz1715.

11. Maupin D, Suchak T, Barnett A, Spick M. Dramatic increases in redundant publications in the Generative AI era. BMC Med. 2026;24:29. doi:10.1186/s12916-025-04569-y.

12. Kusumegi K, Yang X, Ginsparg P, de Vaan M, Stuart T, Yin Y. Scientific production in the era of large language models. Science. 2025;390(6779):1240–3. doi:10.1126/science.adw3000.

13. Volovici V, Syn NL, Ercole A, Zhao JJ, Liu N. Steps to avoid overuse and misuse of machine learning in clinical research. Nat Med. 2022;28(10):1996–9. doi:10.1038/s41591-022-01961-6.

14. Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, Boulesteix AL, Camaradou JC, Celi LA, Denaxas S, Denniston AK, Glocker B, Golub RM, Harvey H, Heinze G, Hoffman MM, Kengne AP, Lam E, Lee N, Loder EW, Maier-Hein L, Mateen BA, McCradden MD, Oakden-Rayner L, Ordish J, Parnell R, Rose S, Singh K, Wynants L, Logullo P. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ. 2024 Apr 16;385:e078378. doi: 10.1136/bmj-2023-078378. Erratum in: BMJ. 2024 Apr 18;385:q902. doi: 10.1136/bmj.q902. PMID: 38626948; PMCID: PMC11019967.

15. Hosseini M, Horbach SPJM. Fighting reviewer fatigue or amplifying bias? Considerations and recommendations for use of ChatGPT and other large language models in scholarly peer review. Res Integr Peer Rev. 2023;8:4. doi:10.1186/s41073-023-00133-5.

16. Laurito W, Davis B, Grietzer P, Gavenčiak T, Böhm A, Kulveit J. AI–AI bias: large language models favor communications generated by large language models. Proc Natl Acad Sci U S A. 2025;122(31):e2415697122. doi:10.1073/pnas.2415697122.

17. Committee on Publication Ethics (COPE). Authorship and AI tools [Internet]. COPE; 2023 [cited 2026 May 28]. Available from: https://publicationethics.org/cope-position-statements/ai-author