Diagn Interv Radiol 2025; DOI: 10.4274/dir.2025.253449

ARTIFICIAL INTELLIGENCE AND INFORMATICS

COMMENTARY

When AI reviews your work: author-centered reflections on LLMs in peer review

Burak Koçak[1]
Mehmet Ruhi Onur[2]

[1]Başakşehir Çam and Sakura City Hospital, Department of Radiology, İstanbul, Türkiye

[2]Hacettepe University Faculty of Medicine, Department of Radiology, Ankara, Türkiye

The integration of large language models (LLMs), such as Chat Generative Pre-trained Transformer (ChatGPT) and Gemini, into peer review has recently emerged as a critical and rapidly evolving issue, raising serious concerns.[1,2] LLMs can be used in various ways during the review process, including language refinement, drafting initial feedback, and even generating full review reports from scratch, yet the extent of their involvement remains unclear.[2,3] Although journals, editors, and reviewers have been the focus of most previous discussions about the use of LLMs in peer review, this commentary shifts attention to authors–the individuals whose unpublished work is being evaluated. Although the core concerns may be shared, authors might experience them from a distinct perspective, shaped by their limited control over the review process and their reliance on it for a fair, expert, and confidential evaluation of their work (Figure 1 and Table 1).

Importantly, and reflecting these very concerns, major academic publishers and journals generally prohibit the use of LLMs in the peer review process, particularly the uploading of manuscripts into such tools.[4,5] However, because these tools are easily accessible, there is a risk that reviewers might use them without disclosure, which would breach editorial policies and bypass oversight. From an author's viewpoint, this potential for unacknowledged LLM use adds another layer of uncertainty to an already non-transparent peer review system.

For authors, one of the most important concerns is the potential breach of confidentiality surrounding their unpublished work.[1,6] The peer review process is conventionally built on a foundation of trust and strict confidentiality, intended to safeguard novel data and ideas from premature or unauthorized disclosure. However, the use of LLMs–particularly general-purpose, widely available models that may store or externally process input–poses a serious risk. If a reviewer inputs all or part of a confidential manuscript into such a model, sensitive content could inadvertently become part of future training data. For authors who have invested substantial time, intellectual effort, and resources into their research, the idea that their findings could be exposed or repurposed before publication is deeply concerning. Even though some LLMs or chat modes claim to offer secure data handling through temporary chat sessions or offline use, authors have no assurance that reviewers will choose or correctly implement these options.

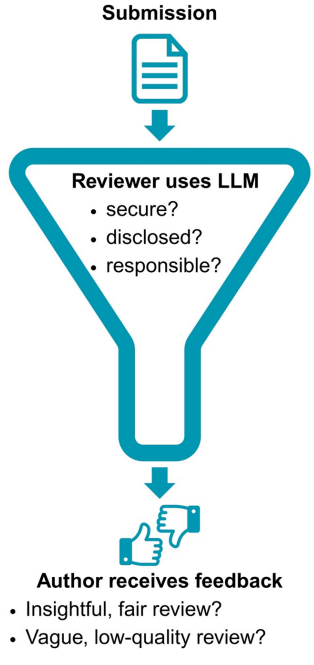

Figure 1. In the context of LLM-involved peer review, the outcome for the author depends on the often unclear and undisclosed extent of LLM use by reviewers. LLM, large language model.

Corresponding author: Burak Koçak

E-mail: drburakkocak@gmail.com

Received 07 May 2025; revision requested 20 May 2025; accepted 25 May 2025.

Epub: 02.06.2025

Publication date:

DOI: 10.4274/dir.2025.253449

You may cite this article as: Koçak B, Onur MR. When AI reviews your work: author-centered reflections on LLMs in peer review. *Diagn Interv Radiol*. 02 June 2025 DOI: 10.4274/dir.2025.253449 [Epub Ahead of Print].

Beyond confidentiality, the quality and reliability of the feedback generated by LLMs pose a major challenge for authors.[7,8] Authors submit their manuscripts expecting insightful, expert critique that helps refine their arguments, methodology, and findings. However, LLMs often lack the nuanced, critical insight of human reviewers.[2] They may generate generic praise or criticism and struggle to evaluate complex or niche academic topics effectively, failing to produce a properly balanced review.[9] Authors receiving such superficial or generic reviews may feel their work has not been truly assessed by an expert, hindering their ability to revise the manuscript effectively. This use of LLMs may result in a notable shift in the author–reviewer dynamic, where authors may develop serious criticisms of the reviewers' reports. Furthermore, authors who focus solely on publishing their work by adhering to reviewer feedback–without questioning its validity–may inadvertently weaken their submission by incorporating misguided or irrelevant revisions, potentially leading to a decline in the quality of the first manuscript draft rather than improvement.

In addition, the potential for inconsistencies, contradictions, and bias in LLM-shaped reviews can create confusion and frustration for authors.[10,11] LLMs are highly sensitive to prompt variations, meaning that even slight changes in phrasing can produce markedly different responses. This variability may lead to internally inconsistent reviews or comments that contradict feedback from other reviewers. LLMs can also exhibit sycophancy, aligning with a reviewer's biased phrasing rather than the manuscript's objective content. From an author's perspective, receiving contradictory or unclear feedback makes it difficult to identify valid points for revision. Compounding this, LLMs may demonstrate bias–potentially favoring papers from well-known authors or prestigious institutions if the review is not blinded.[11] This raises concerns about fairness and equity in the evaluation process, particularly for authors from less prominent backgrounds.

Another notable concern is the tendency of LLMs to generate irrelevant or fabricated content, including fictitious references.[9] Authors may receive comments based on non-existent issues or be asked to address points supported by fabricated citations. Identifying these "hallucinations" requires authors–or editors–to critically scrutinize every detail of the review, adding another layer of burden to the already demanding process of manuscript revision.

Perhaps the most fundamental problem from the author's viewpoint is the lack of transparency regarding LLM use in review.[12] Reviewers may not disclose their use of AI tools, and the inherent opacity of LLMs–combined with tools designed to make AI-generated text appear human-like–makes detection challenging for editorial teams. This means authors may receive a review shaped or even generated by an LLM without knowing it. Without this knowledge, authors are ill-equipped to interpret the feedback appropriately or to advocate for their work in response to potential LLM idiosyncrasies such as hallucinations or contradictions.

Recognizing these challenges, one key recommendation is to notify authors if the peer review process involves LLM assistance.[2] This disclosure is crucial, as it allows authors to understand the potential influence of the tool and to respond accordingly to feedback that may reflect LLM limitations. It empowers authors to critically evaluate the review and address possible flaws attributable to AI rath-

Figure 2. Author responsibilities and recommended actions in the context of LLM-involved peer review. Although individual measures are generally feasible, community-level recommendations require collective action, such as by editorial boards or professional societies. LLM, large language model.

Table 1. Key author-centered concerns regarding LLM-involved peer review and their implications

Concern	Description	Implications for authors
Confidentiality	Reviewer may input the manuscript content into general-purpose LLMs	Potential breach of confidentiality; unauthorized reuse of unpublished ideas or data
Feedback quality	Feedback may be overly generic, superficial, or context-insensitive	Limited value in improving the manuscript; lack of expert-level critique; the possibility of misguidance
Hallucination risk	LLMs may introduce fictitious references or identify non-existent flaws	Authors may waste effort addressing invalid or fabricated concerns
Inconsistency	Responses may be internally inconsistent or conflict with other reviewers' comments	Challenges in interpreting and responding to contradictory or incoherent feedback
Bias and manipulation	LLMs may favor prestigious authors or verbose texts; vulnerable to prompt manipulation	Risk of unfair assessments and unintentional reinforcement of systemic biases
Lack of transparency	Reviewers may not disclose their use of LLMs	Authors may be unaware of AI-generated content and unprepared to interpret LLM-specific issues. Familiarity with AI-generated generic, irrelevant, or fabricated language may erode trust in reviewers and the integrity of journal reviews

LLM, large language model; AI, artificial intelligence.

er than blindly accepting potentially inaccurate or irrelevant comments.

In conclusion, while often unspoken, LLM involvement in peer review may be more common than acknowledged and is likely to increase with the widespread availability of these tools. For authors, the integrity of peer review depends on receiving expert, objective, reliable, and confidential evaluations. Rather than pursuing an unrealistic ban, the focus should shift toward managing LLM use responsibly–ensuring strong human oversight and critical judgment so that LLMs support, rather than undermine, the peer review process.[13] Safeguarding the integrity of peer review requires clear journal policies, targeted training for editors and reviewers, and transparency with authors to enable informed responses. Authors, both as contributors and community members, play a critical role in upholding peer review standards amid increasing LLM involvement. They, in turn, should remain vigilant and adopt best practices to protect the integrity of their work (Figure 2).

Footnotes

Acknowledgments

Language of this manuscript was checked and improved by ChatGPT (4o). The authors conducted strict supervision when using this tool.

Conflict of interest disclosure

Burak Koçak, MD, serves as Section Editor for Diagnostic and Interventional Radiology (DIR). Mehmet Ruhi Onur, MD, is Editor-in-Chief of DIR. They had no involvement in the peer review of this article and had no access to information regarding its peer review.

References

1. 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(1):4. [Crossref]
2. Kocak B, Onur MR, Park SH, Baltzer P, Dietzel M. Ensuring peer review integrity in the era of large language models: a critical stocktaking of challenges, red flags, and recommendations. *European Journal of Radiology Artificial Intelligence*. 2025;2:100018. [Crossref]
3. Zhou L, Zhang R, Dai X, Hershcovich D, Li H. Large language models penetration in scholarly writing and peer review. Published online February 16, 2025. [Crossref]
4. Hamm B, Marti-Bonmati L, Sardanelli F. ESR Journals editors' joint statement on Guidelines for the use of large language models by authors, reviewers, and editors. *Insights Imaging*. 2024;15(1):18. [Crossref]
5. Moy L. Guidelines for use of large language models by authors, reviewers, and editors: considerations for imaging journals. *Radiology*. 2023;309(1):e239024. [Crossref]
6. Zhuang Z, Chen J, Xu H, Jiang Y, Lin J. Large language models for automated scholarly paper review: a survey. Published online January 17, 2025. [Crossref]
7. Liang W, Zhang Y, Cao H, et al. Can large language models provide useful feedback on research papers? A large-scale empirical analysis. *NEJM AI*. 2024;1(8):AIoa2400196. [Crossref]
8. Ou J, Walden WG, Sanders K, et al. CLAIMCHECK: how grounded are LLM critiques of scientific papers? Published online March 27, 2025. [Crossref]
9. Donker T. The dangers of using large language models for peer review. *Lancet Infect Dis*. 2023;23(7):781. [Crossref]
10. Borji A. A Categorical archive of ChatGPT failures. Published online April 3, 2023. [Crossref]
11. von Wedel D, Schmitt RA, Thiele M, Leuner R, Shay D, Redaelli S, Schaefer MS. Affiliation bias in peer review of abstracts by a large language model. *JAMA*. 2024;331(3):252-253. [Crossref]
12. 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(8):702-703. [Crossref]
13. Ebadi S, Nejadghanbar H, Salman AR, Khosravi H. Exploring the impact of generative AI on peer review: insights from journal reviewers. *J Acad Ethics*. Published online February 11, 2025. [Crossref]