16 February 2026 · CC-BY 4.0 Qeios

Peer Review

Review of: "The Role of Artificial Intelligence in the Lifecycle of Scientific Manuscripts: Authoring, Reviewing, and Editorial Selection"

Manuel Rey-Álvarez Zafiria[1]

1. Universidad Autónoma de Madrid, Spain

NOTE: *The corrections suggested below are based on version 1 of the publication. I have reviewed the second version of the manuscript and acknowledge the author's efforts to clarify the genre (now positioned as critical commentary) and to introduce nuances regarding non-commercial journals and legal responsibility. These improvements are welcome. However, the substantial technical and economic shortcomings identified in my evaluation of the earlier version remain entirely unaddressed.*

REVIEW:

I agree with this publication in its essentials. However, the manuscript suffers from the following problems:

1. False dichotomy: A simplistic opposition is frequently presented between "technical but biased AI" and the "creative and ethical human." Reality is more complex. Humans are also deeply biased, inconsistent, and often not very creative in our reviews. AI could, if properly designed, help correct certain human biases (e.g., affiliation bias). The author mentions this in passing but does not develop it.
2. Scenario 3 in Table 3 is desirable, but the author does not specify how to implement it technically. What type of AI? At what level of transparency (explainable AI)? How would human feedback be integrated to improve the models? Without these details, the proposal remains an ethical desideratum.
3. The author builds much of the argument on the "existential threat" of hallucinations and the "inherent incapacity" of AI for tasks such as bibliographic verification or novelty assessment.

 doi.org/10.32388/0FYAK0

However, an entire set of techniques and architectures developed precisely to address these problems is completely ignored:

1. Chain-of-Thought (CoT) and extended reasoning: The ability of models to "think step by step" before responding is not a mere embellishment; it represents a paradigm shift in reliability. CoT enables the model to decompose complex problems (such as verifying a citation chain or assessing the internal coherence of a statistical method) into subproblems, drastically reducing reasoning errors. A model using CoT does not merely "predict the next word," but simulates a deliberative process.
2. Internet search and real-time information retrieval (RAG 2.0): The author mentions RAG in passing and quickly dismisses it as insufficient ("cannot substitute for a true understanding"). However, modern RAG systems are not mere appendices; they form the backbone of research assistants such as Elicit, Scite, or Consensus. These systems do not "hallucinate" citations because they do not generate them: they retrieve them from indexed databases (PubMed, arXiv, Crossref) and present them with verifiable links. A model that, when asked about a paper, searches Google Scholar in real time and extracts information from the source is not subject to the same type of hallucination as a "closed" LLM. Domingo treats all AI systems as if they were ChatGPT 3.5 without internet access, which is technically obsolete.
3. Automated reference verification: There are tools (some based on LLMs, others on traditional search engines) that can automatically cross-check each generated citation against databases such as CrossRef or Scopus, flagging discrepancies. The author mentions the problem of incorrect DOIs but does not discuss emerging solutions: systems that not only detect that a DOI is false but attempt to identify the correct DOI of the paper intended to be cited, based on context.
4. Factual consistency models: AI research has for years been developing specific metrics and models to measure the factual consistency of a generated text with respect to a given source. The point is not that the LLM "understands" truth, but that it is possible to construct a system that verifies whether what the LLM states is supported by the sources provided to it.

The author presents risks (hallucinations, biases) as existential and immutable failures of the technology, rather than as engineering problems that are being actively addressed. This seems to me a missed opportunity: the article could have explored how the technical community is responding to these challenges, instead of portraying the technology as a static and dangerous monolith.

Having addressed the technical dimension, I would like to offer an "economic" review:

The author rightly denounces the "exploitation" of reviewers' unpaid labor, but the proposed solution is presented without even a minimally realistic analysis of its economic feasibility. In Section 5.1, a stipend of €200–€500 per review is suggested. Let us perform some rough calculations to understand the magnitude of the problem:

Volume of manuscripts: According to 2023 data, approximately 3 million academic articles were published in peer-reviewed journals. To this must be added rejected manuscripts, which in many high-impact journals exceed 80–90% of submissions. A conservative estimate of total submissions (manuscripts requiring review) could be between 8 and 10 million per year.

Number of reviews per manuscript: The average is typically 2 to 3 reviews per manuscript (considering successive rounds). Let us take 2.5 as a weighted average.

Calculation of total cost:

If we take the lower end of Domingo's proposal (€200):
Total reviews per year = 9 million submissions × 2.5 reviews = 22.5 million reviews.
Total cost = 22.5M × €200 = €4.5 billion annually.
If we take the upper end (€500):
Total cost = 22.5M × €500 = €11.25 billion annually.

We are speaking of an industry that, according to estimates, generates between $10 billion and $19 billion annually in revenue from APCs and subscriptions. Paying reviewers would, at best, increase industry costs by 30%, and at worst, double them or more.

The problem here is cost transfer, because someone would undoubtedly have to pay this additional expense—but who?

The author seems to forget that the current system of "unpaid labor" is not free; it carries an enormous opportunity cost borne by researchers and their institutions. Paying €200 per review does not create new value out of nothing; it merely transfers the cost from the researcher (in time) to the publisher (in money). But that money must come from somewhere.

doi.org/10.32388/0FYAK0

If institutions (universities, research centers) are ultimately the ones paying APCs (directly or through library consortia), Domingo's proposal becomes a closed loop of self-financing:

A simple example:

The university pays an APC of €3,000.

The publisher, with that money, pays €200 to the reviewer (who works at the same university).

The net result is that the university has paid €3,000 and recovered (in cash to its researchers) only a very small fraction. The rest is retained by the publisher or finances the process.

This does not resolve systemic "exploitation"; it merely monetizes it in an inefficient and likely regressive manner (wealthy universities will be able to pay high APCs and their researchers will "recover" something through reviews; poorer universities will not). Moreover, it creates an incentive for publishers to respond to regulation by massively increasing costs to secure their margins.

I believe there are two solutions the author does not consider and which would be worth analyzing in the publication, in relation to the aforementioned Deep Learning architectures that improve reasoning, reliability, and reduce hallucinations in models:

a. Diamond models (Diamond Open Access): Journals managed by academic communities or universities, non-profit, that do not charge APCs. Here, "paying" the reviewer is impossible because there is no cash flow. Sustainability is achieved through institutional labor (librarians, university staff) and voluntarism, precisely what the author criticizes. The difference is that there is no shareholder profiting in the middle. In some respects, it resembles the philosophy of open source software. I firmly believe this is the model the scientific community should promote.
b. Efficiency via AI (properly understood): AI, as discussed above, can drastically reduce the time a human needs to devote to a review. If an AI verifies citations, formatting, statistics, and plausibility in five minutes, the human reviewer can focus on what is substantive: the argument. One hour instead of three.

This reduces the reviewer's opportunity cost. It makes the system more sustainable without the need for massive direct payment (although a symbolic payment or accessible subscription models could still be justified).

Open Peer Review: The fact that the article has 8 open reviews on the platform itself is in its favor. It addresses a complex and highly topical issue, and I believe the author has been very courageous in tackling it. The review by Andrew Williams, which calls for a more nuanced exposition, and that of Osamah Mohammed Alyasiri, which questions the genre of the article, are particularly incisive and should be incorporated by the author in a future version.

Should this work be published? Yes, but with substantial modifications.

Declarations

Potential competing interests: No potential competing interests to declare.