



EDITORIAL

Rethinking value in science: Moving beyond the stigma of artificial intelligence in academic communication

Repensando el valor en la ciencia: más allá del estigma de la inteligencia artificial en comunicación académica

Carlos González Morcillo

Escuela Superior de Informática, Universidad de Castilla-La Mancha, Ciudad Real, Spain

The use of artificial intelligence (AI) in the different stages of scientific publication has spurred an intense debate in the academic community. Some top-notch journals, such as *Science* or *Nature*, have banned the use of large language models (LLMs) citing ethical concerns regarding authorship of the texts.¹ However, a large part of the arguments in this debate are based on obsolete premises that must be urgently reconsidered. Just as it would be absurd to penalize a researcher for using statistical analysis software instead of carrying out the analyses by hand, does it make sense to question the use of AI when it is not but yet another tool to optimize scholarly communication?

The fact of the matter is that we are facing a revolution in how science is communicated. Large language models (such as ChatGPT, Claude, Llama or Mixtral, among others) currently allow the easy generation of texts that are flawless in terms of grammar and structure.³ The scientific quality of a publication must take these criteria for granted, which is directly fulfilled with these tools, and its assessment should rather focus on the four key criteria: (1) originality, (2) methodological rigor, (3) relevance and (4) impact of the scientific content, which can hardly be automated.² These

models become essential allies for researchers whose first language is not English, which continues to be the dominant language in the *JCR* top-ranking journals.

Large language models are now in the position previously held by statistical machine translation models, such as DeepL, or human translators who, despite their language abilities, frequently were unfamiliar with the specific terminology and nuances of each specific field.¹

Lately, there has been a proliferation of guidelines and recommendations that address the use of AI in the authorship, peer review and editorial processing of scientific manuscripts.⁴ Many suggest explicitly reporting the use of these models in the manuscript writing process, either by listing them as a coauthors or stating their use in the acknowledgment section, specifying the model used, its version and the date of use. However, these recommendations are groundless. Large language models are tools that generate text sequentially, in response to the prompts and context provided by the user. This controlled nature entails that an LLM cannot be considered responsible for its output, as it does not generate text except in response to a specific prompt.¹

Disclosing the specific use of an LLM is irrelevant on account of the stochastic nature of these models. Due to the temperature parameter, which controls the randomness in the generation of each token (tokens are sequentially generated pieces of words or text), the output of the model

DOI of original article:

<https://doi.org/10.1016/j.anpedi.2024.503696>

E-mail address: carlos.gonzalez@uclm.es

may vary even with identical prompts. In consequence, even using the same version of the model, the reproducibility of the generated text cannot be guaranteed. If we consider LLMs as advanced software tools, then they do not merit special mention in acknowledgments or any other credits. We do not acknowledge or express thanks for the use of Google, the Web of Science, the R programming language or SPSS. Why make an exception with this technology? Furthermore, what sense would it make once these models become integrated in word-processing applications, such as Word or Overleaf? We never applied this preferential treatment to the controversial Clippy assistant in Microsoft Word.

The effective detection of the use of LLMs in text generation is technically complex, especially when techniques are used to modify the vocabulary of the model. In fact, such modifications may make it nearly impossible to determine whether a text has been produced with the assistance of AI, which further challenges the usefulness of policies focused on such detection rather than the quality of the content.

Paradoxically, while we use sophisticated statistical and technological tools in our research, we continue holding on to a romantic and outdated view of the scientific writing process. This is reminiscent of Goodhart's law, as applied to the academic environment: when the way in which something is written becomes a target in itself, we cease to assess what is truly important.⁵ Just as academic qualifications do not reliably predict professional performance, quality in research is not founded on the accumulation of publications in journals in the first and second JCR quartiles (enter your list of top *predatory* journals here). When we use an indicator to make decisions or assess performance, people start to optimize their activity with the aim of maximizing that indicator. In the process, the "100% human writing" indicator loses its value as a measure, because it is manipulated instead of promoting its original objective: the production of high-quality scientific content.

Scientific journals need to shift toward a new paradigm by which the actual content of contributions will be assessed, rather than the means used to convey them. This means that we have to rethink our editorial processes at 3 fundamental levels, each of which poses its own challenges and opportunities.

As concerns authors, we must stop thinking of AI as a "cheating" tool and recognize it as what it actually is: a legitimate assistant that makes it possible to focus on what truly matters, which is research. The effective use of AI will become the norm, similar to the use of advanced statistical software in data analysis. However, there are potential risks to the use of LLMs, especially in highly sensitive fields, such as medical research.³ Articles in this field contain a lot of contextual information, with critical nuances and subtle details that require special attention. In addition, the intrinsic bias in these models calls for a meticulous human analysis to guarantee the integrity and accuracy of the data.

As concerns peer reviewers, AI can become a powerful ally, allowing them to focus on the assessment of the essential scientific content and freeing them from the tedious task of revising formal aspects. A recent study² that analyzed more than 3000 articles from scientific conferences demonstrated that AI tools achieve a performance comparable to that of humans in the preliminary review stage of the peer review process. However, in later stages, where key criteria such as originality, significance and rigor are assessed, the potential impact of AI turned out to be significantly lower.

In the context of editorial and review work, there is a critical concern that is often overlooked: the confidentiality of data. When cloud-based LLMs are used (as opposed to open source or locally run models), the contents of manuscripts are uploaded to external platforms that could use these data in the training of future versions of the model. This raises significant questions regarding the handling of unpublished scholarly content and the responsibility of editors and reviewers in protecting the intellectual property of authors.

Editors face the challenge of redefining the metrics used to assess scientific output. The extent to which LLMs facilitate rewriting not only renders traditional means for the detection of plagiarism useless, but also demands that the value of the scientific content is prioritized over formal concerns.

Institutions that ban the use of AI are committing a strategic blunder, much like institutions in the past that resisted the use of the Internet or electronic databases. They are depriving their researchers of valuable tools that could significantly improve the quality and scope of their work. The solution is not to ban the use of generative algorithms, but to teach how to use them effectively and responsibly. Artificial intelligence does not pose a threat to academic integrity; it is a chance to democratize and improve scientific communication at a global level. And it is certainly not cheating.

References

1. Hosseini M, Resnik DB, Holmes K. The ethics of disclosing the use of artificial intelligence tools in writing scholarly manuscripts. *Res Ethics*. 2023;19:449–65.
2. Checco A, Bracciale L, Loreti P, Pinfield S, Bianchi G. AI-assisted peer review. *Humanit Soc Sci Commun*. 2021;8:1–11.
3. Hryciw BN, Seely AJ, Kyeremanteng K. Guiding principles and proposed classification system for the responsible adoption of artificial intelligence in scientific writing in medicine. *Front Artif Intell*. 2023;6:1283353.
4. Leung TI, de Azevedo Cardoso T, Mavragani A, Eysenbach G. Best practices for using AI tools as an author, peer reviewer, or editor. *J Med Internet Res*. 2023;25:e51584.
5. Kaebnick GE, Magnus DC, Kao A, Hosseini M, Resnik D, Dubljević V, et al. Editors' statement on the responsible use of generative AI technologies in scholarly journal publishing. *Med Health Care Philos*. 2023;26:499–503.