COMMENTARY

Generative AI for Academic Publishing? Some Thoughts About Epistemic Diversity and the Pursuit of Truth

Lai Ma

University College Dublin

The uses of generative AI have prompted both positive and negative responses. This short commentary contemplates potential issues concerning bibliodiversity, epistemic diversity, and data surveillance. It also cautions the potential erosion of public trust in academic publishing in the age of generative AI. Invoking the Sokal hoax, the commentary sheds light on what it means when knowledge, experience, expertise, and the pursuit of truth are on the line.

Keywords: generative AI; scholarly communication; academic publishing; epistemic diversity; bibliodiversity; machine learning; large language models; biases

Introduction

In 1996, Alan Sokal, a physics professor, published an article entitled "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity" in the "Science Wars" issue of *Social Text,* a cultural studies journal. The article includes fifty-five lengthy endnotes, with references to a wide array of authors from David Bloor, Jacques Derrida, and Sandra Harding to Jean-François Lyotard, Immanuel Wallerstein, and Slavoj Žižek. One paragraph reads:

> Second, the postmodern sciences deconstruct and transcend the Cartesian metaphysical distinctions between humankind and Nature, observer and observed, Subject and Object. Already quantum mechanics, earlier in this century, shattered the ingenuous Newtonian faith in an objective, prelinguistic world of material objects "out there"; no longer could we ask, as Heisenberg put it, whether "particles exist in space and time objectively." But Heisenberg's formulation still presupposes the objective existence of space and time as the neutral, unproblematic arena in which quantized particle-waves interact (albeit indeterministically); and it is precisely this would-be arena that quantum gravity problematizes. Just as quantum mechanics informs us that the position and momentum of a particle are brought into being only by the act of observation, so quantum gravity informs us that space and time themselves are contextual, their meaning defined only relative to the mode of observation. (Sokal [1996] 2000a, 22)

Then, in the May–June 1996 issue of *Lingua Franca,* Sokal revealed that the "Transgressing the Boundaries" article was a hoax (Sokal [1996] 2000b). Self-identified as a political leftist, he was concerned about the denial of objective or scientific truths in the postmodernist discourse after reading *Higher Superstition: The Academic Left and Its Quarrels with Science* (Gross and Levitt 1994).

One can disagree with Sokal's arguments or tactic, but there is no denying that the "Transgressing the Boundaries" article was not easy to write. The piece must have taken up substantial time and energy to compile and compose—and Sokal could not have known that it would actually be published. The publication of the article, if nothing else, demonstrated that sometimes bogus articles could sail through the editorial and peer review process. The Sokal hoax was an intervention that generated thought-provoking discussions about knowledge and epistemology, scientific facts and cultural phenomena. The hoax was a big deal.

Today, many are seemingly unaware of the volume of articles published by predatory journals (or journals of questionable practices) and produced by paper mills. Even high-profile retractions due to fraudulent research practices such as fabrication of data and images (Van Noorden 2023) do not seem to trigger fervent discussions about research integrity. While we have PubPeer and Retraction Watch, there have been few oppositions and genuine discussions as to what should be committed to the scholarly record. One wonders, would a twenty-first-century Sokal hoax elicit meaningful responses from the scholarly community? Or perhaps it has been accepted that bogus articles are the by-products of academic publishing nowadays? And what will the research community be willing to accept when seemingly nonsensical articles can be composed by generative AI, when every article can look like a Sokal hoax, when one needs no knowledge of either hermeneutics or quantum physics, but just a few clever prompts, to churn out an article like "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity"?

The Scopus Corpus

Elsevier recently announced the launch of Scopus AI, a search tool that can produce overviews of research topics and sometimes even highlight gaps in the literature (Evans 2024; Grove 2024). It is not possible to prompt Scopus AI to produce an article about quantum gravity using postmodernist theories (yet), but Scopus AI seems promising: the growth of articles is certain because researchers can now produce more articles in less time. One advantage of Scopus AI, compared to Open AI's ChatGPT or Alphabet's Gemini (previously Bard), is that it "learns" from published works in Scopus-indexed journals (hereafter "Scopus corpus"). Put another way, the Scopus corpus consists of legitimate and reputable sources that have passed through peer reviewers and editorial boards, meaning that the summaries Scopus AI generates are from presumably reliable and trusted sources. Introducing the new search tool, an article on Elsevier Connect states that

> Scopus AI draws from metadata and abstracts of Scopus documents published since 2013. Advanced prompt engineering and curated recent data minimize risks of false AI-generated information and ensure responses are based on recent, trusted knowledge. (Evans 2024)

True, the quality assurance about "recent, trusted knowledge" is important, yet it is necessary to acknowledge the limited coverage and biases of the Scopus corpus—in particular, how they will exacerbate epistemic injustice in the age of AI academic publishing. First, the predominance of English-language journals published in North America and Western Europe means that the summaries and concept maps generated by Scopus AI will reproduce contents that are of primarily Western focus, especially by researchers affiliated with well-resourced institutions. This is, of course, not to discount the quality or significance of their work. However, it is very likely that Scopus AI lacks knowledge and training in non-Western research topics and perspectives due to the limited scope of the Scopus corpus. While the search results can be useful in many cases, they can also be restrictive and limited in prompting innovative and novel ideas and approaches.

Second, when Scopus AI generates a list of references, it is not searching beyond the Scopus corpus. Since it is very likely that these lists of references will become citations in research articles, it will automatically enhance the citation advantage of the Scopus corpus. In other words, Scopus AI will reinforce the legitimacy and the presumed quality of Scopus-indexed publications. Together, the Scopus AI summaries and references lists are bound to perpetuate the privilege of Scopus-indexed publications when researchers become dependent on the tool and less inclined to search for and read publications elsewhere.

In other words, the tool is not conducive to bibliodiversity. Even without Scopus AI, many have voiced concerns about the presumed "international" standard of Scopus-indexed publications (Beigel 2021; Mills et al. 2021). Scopus AI and similar tools can be useful and effective, but they can further propagate monoculture (see, for example, Demeter and Toth 2020) and obstruct the growth of bibliodiversity. Since publications and citations are still held as important tokens in researchers' careers and university rankings, Scopus AI could mean that publishing in Scopus-indexed publications would become ever more important because one's work will definitely not be learned by the Scopus AI machine, and hence not included in their lists of references, if one is publishing elsewhere (i.e., publications not indexed by Scopus).

There is another danger inherent in the Scopus corpus (or any corpus) for machine learning: as the saying goes, "Garbage in, garbage out." Generative AI is powered by machine learning using large language models (LLMs). The outputs of generative AI are largely dependent on the inputs fed into LLMs. Generally speaking, the higher quality the inputs, the better the outputs. In scholarly publishing, the drastic increase of publications, the peer review crisis, and the instances of research misconduct mean that even a corpus of "trusted knowledge" needs to be critically examined. It is one thing to cede reading to machines (Carpenter 2024);

it is another thing to trust AI-generated summaries that may not be able to make distinctions between exploratory and large-scale studies, inconclusive and generalizable results. Biases and falsehoods in machine learning stem from the corpora used to train the algorithms. These biases and falsehoods cannot be ignored, especially when Scopus AI possibly means an increase in the volume and pace of scholarship; as more publications are produced at a faster pace with AI, those publications risk further perpetuating the errors and biases in the original corpus.

Acceleration, Accumulation

A survey conducted by *Nature* in 2023 gauged researchers' views of AI. In the responses, "faster" and "speed" were highlighted as the major positive impacts and benefits (Table 1).

The results raise interesting questions as to what it means when AI-powered search tools replace reading, critical thinking, and generating syntheses and research questions. But more importantly, the results raise concerns about the platformization of scholarly information (Ma 2023a, 2023b) and surveillance publishing (Pooley 2022, 2024). It is because a tool such as Scopus AI entails harvesting, packaging, and selling data in many forms and formats, including citation data, collaboration networks, and data about researchers and their research and other online activities. These practices and threats already exist (Yoose and Shockey 2024), and they will intensify when machine learning becomes more pervasive and persistent. As AI tools make the publication process easier and faster, there will be more data traffic to be captured and possibly sold. The speed of the hamster wheel is going to accelerate.

Furthermore, epistemic diversity and bibliodiversity will be more challenging to achieve if Scopus AI or similar tools take hold. It is because diversity takes time to develop. Bibliodiversity also means nurturing and recognizing materials that are not in the Scopus corpus. When these materials are not ingested by the AI tools and are excluded from the summaries and lists of references generated, the neglect and ignorance will only continue and possibly worsen. Scopus AI is for acceleration and capital accumulation: those who can afford the tool will benefit from producing more publications and accumulating more citations.

Table 1: Positive impacts and benefits of generative AI presented in "AI and Science: What 1,600 Researchers Think" (Van Noorden and Perkel 2023)

Positive impacts of AI	Benefits of generative AI
Provides faster ways to process data	Helps researchers without English as a first language (through editing or translation)
Speeds up computations	Makes coding easier and faster
Saves researchers time or money	Summarizes other research to save time reading it
Automates data acquisition	Speeds administrative tasks
Makes it possible to process new kinds of data	Helps write manuscripts faster
Provides faster ways to write code	Improves scientific research
Answers questions that are otherwise very difficult to solve	Helps creative work by brainstorming new ideas
Optimizes experimental set-ups for acquiring data	Generates new research hypotheses
Makes new discoveries	Helps peer-review manuscripts faster
Generates new research hypotheses	

Conclusion

One might regard Sokal's approach with distaste and one might not agree with his reading of postmodernism and/or specific texts, but one cannot deny that Sokal acted because of his concerns for and interests in the pursuit of truth. There was no guarantee that his submission would be accepted for publication. The writing of the article was not assisted by ChatGPT or Scopus AI. It was a hoax that took serious work—and some understanding (if inaccurate or superficial) of the cited works. Whether or not the article or the lengthy notes actually make sense, they took time to compile and write.

When the hoax was revealed, researchers actively engaged in ensuing discussions and debates. This is not to say that there were no egoistic warriors or epistemic partisans. Surely there were disagreements, and even anger, but it was largely because the pursuit of knowledge and truth was on the line. There was space for epistemic diversity. The Sokal hoax was an affair that took away time and energy from applying for grants, doing research, or writing publications. Fast forward to 2024, it seems that even the alarming number of retractions and reports of research misconduct (Van Noorden 2023) do not trigger much reaction and reflection, much less "sh[ake] the academy" (Editors of *Lingua Franca* 2000). It raises questions as to how many retractions and how much misconduct will be tolerated when the academic marketplace is urging for more publications in faster cycles and when more and faster can be assisted by Scopus AI?

Scopus AI is a form of automation, and much work has been written about the relationship between automation and inequality (see Eubanks 2017). While automation is nothing new, it seems that every new wave of automation has one thing in common: a small proportion of people becomes richer (or extremely rich) while most people live and work under more stress with less time for family, friends, and leisure. Writing about the original Luddites, Merchant (2023) has aptly pointed out that the rebellion was less about new technologies and more about changes in community, social structure, and the distribution of wealth. The ownership of new technologies means accumulation of capital; those without may suffer from unemployment or exploitative labour practices. In the context of academic publishing, Scopus AI will not alleviate the Matthew effect (Merton 1968)—that is, the disproportionate credit given to established researchers according to the principle of cumulative advantage. In fact, it will potentially make the distribution more extreme. The data cartels and their monopoly of scholarly information (Lamdan 2023) will continue, while bibliodiversity will struggle to survive and thrive.

"The Emperor has no clothes" is used as an expression by some commentators of the Sokal hoax (Editors of *Lingua Franca* 2000). A long-term consequence of the use of generative AI in academic publishing without responsible and sufficient safeguards may be the loss of public trust, for scholarly works could become the emperor's new clothes. At the time of this writing, an article using AI-generated images is being widely reported in social and mass media (Pearson 2024). It is an alarming example of what AI will bring when academic publishing is about acceleration, accumulation, and profit-maximising, when speeding up causes harm to quality, credibility, and trust. The implications of automation in knowledge production demand careful consideration, as knowledge, experience, and expertise—and the pursuit of truth—can be undermined in the age of generative AI.

Competing Interests

The author declares no conflict of interest.

References

Beigel, Fernanda. 2021. "A Multi-Scale Perspective for Assessing Publishing Circuits in Non-Hegemonic Countries." *Tapuya: Latin American Science, Technology and Society* 4 (1): 1–16. https://doi.org/10.1080/25729861.2020.1845923.

Berger, Monica. 2021. "Bibliodiversity at the Centre: Decolonizing Open Access." *Development and Change* 52 (2): 383–404. https://doi.org/10.1111/dech.12634.

Carpenter, Todd A. 2024. "Let's Be Cautious as We Cede Reading to Machines." *The Scholarly Kitchen*, January 25, 2024. https://scholarlykitchen.sspnet.org/2024/01/25/lets-be-cautious-as-we-cede-reading-to-machines. Archived at: https://perma.cc/NT5N-ZLPS.

Demeter, Marton, and Tamas Toth. 2020. "The World-Systemic Network of Global Elite Sociology: The Western Male Monoculture at Faculties of the Top One-Hundred Sociology Departments of the World." *Scientometrics* 124 (3): 2469–95. https://doi.org/10.1007/s11192-020-03563-w.

Editors of *Lingua Franca*, eds. 2000. *The Sokal Hoax: The Sham That Shook the Academy*. Lincoln: University of Nebraska Press.

Eubanks, Virginia. 2017. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.

Evans, Ian. 2024. "How Researchers Can Use GenAI to Get the Info They Need Quicker Than Ever." *Elsevier Connect*, February 15, 2024. https://www.elsevier.com/connect/how-researchers-can-use-genai-to-get-the-info-they-need-quicker-than-ever. Archived at: https://perma.cc/XB47-VMZB.

Gross, Paul R., and Norman Levitt. 1994. *Higher Superstition: The Academic Left and Its Quarrels with Science*. Baltimore: Johns Hopkins University Press.

Grove, Jack. 2024. "Elsevier Launches Scopus AI Bot for Literature Reviews." *Times Higher Education*, January 16, 2024. https://www.timeshighereducation.com/news/elsevier-launches-scopus-ai-bot-literature-reviews. Archived at: https://perma.cc/Y3CR-QRMM.

Lamdan, Sarah. 2023. *Data Cartels: The Companies That Control and Monopolize Our Information*. Redwood City: Stanford University Press.

Ma, Lai. 2023a. "Information, Platformized." *Journal of the Association for Information Science and Technology* 74 (2): 273–82. https://doi.org/10.1002/asi.24713.

Ma, Lai. 2023b. "The Platformisation of Scholarly Information and How to Fight It." *LIBER Quarterly: The Journal of the Association of European Research Libraries* 33 (1): 1–20. https://doi.org/10.53377/lq.13561.

Merchant, Brian. 2023. *Blood in the Machine: The Origins of the Rebellion Against Big Tech*. New York: Little, Brown and Company.

Merton, Robert K. 1968. "The Matthew Effect in Science: The Reward and Communication Systems of Science Are Considered." *Science* 159 (3810): 56–63. https://doi.org/10.1126/science.159.3810.56.

Mills, David, Abigail Branford, Kelsey Inouye, Natasha Robinson, and Patricia Kingori. 2021. "'Fake' Journals and the Fragility of Authenticity: Citation Indexes, 'Predatory' Publishing, and the African Research Ecosystem." *Journal of African Cultural Studies* 33 (3): 276–96. https://doi.org/10.1080/13696815.2020.1864304.

Pearson, Jordan. 2024. "Scientific Journal Publishes AI-Generated Rat with Gigantic Penis in Worrying Incident." *Vice*, February 15, 2024. https://www.vice.com/en/article/dy3jbz/scientific-journal-frontiers-publishes-ai-generated-rat-with-gigantic-penis-in-worrying-incident. Archived at: https://perma.cc/U4BM-432P.

Pooley, Jeff. 2022. "Surveillance Publishing." *The Journal of Electronic Publishing* 25 (1). https://doi.org/10.3998/jep.1874.

Pooley, Jeff. 2024. "Large Language Publishing." *Upstream*, January 2, 2024. https://doi.org/10.54900/zg929-e9595.

Sokal, Alan. (1996) 2000a. "Transgressing the Boundaries: Toward a Transformative Hermeneutics of Quantum Gravity." In *The Sokal Hoax: The Sham That Shook the Academy*. Edited by the editors of *Lingua Franca*, 11–45. Lincoln: University of Nebraska Press.

Sokal, Alan. (1996) 2000b. "Revelation: A Physicist Experiments with Cultural Studies." In *The Sokal Hoax: The Sham That Shook the Academy*. Edited by the editors of *Lingua Franca*, 49–53. Lincoln: University of Nebraska Press.

Van Noorden, Richard. 2023. "More Than 10,000 Research Papers Were Retracted in 2023 — a New Record." *Nature* 624: 479–81. https://doi.org/10.1038/d41586-023-03974-8.

Van Noorden, Richard, and Jeffrey M. Perkel. 2023. "AI and Science: What 1,600 Researchers Think." *Nature* 621: 672–75. https://doi.org/10.1038/d41586-023-02980-0.

Yoose, Becky, and Nick Shockey. 2024. "Navigating Risk in Vendor Data Privacy Practices: An Analysis of Elsevier's ScienceDirect." SPARC. https://doi.org/10.5281/zenodo.10078610.