

# Reimagining Scholarship: A Response to the Ethical Concerns of AUTOGEN

Hazem Zohny

**To cite this article:** Hazem Zohny (2023) Reimagining Scholarship: A Response to the Ethical Concerns of AUTOGEN, The American Journal of Bioethics, 23:10, 96-99, DOI: [10.1080/15265161.2023.2250315](https://doi.org/10.1080/15265161.2023.2250315)

**To link to this article:** <https://doi.org/10.1080/15265161.2023.2250315>



© 2023 Taylor & Francis Group, LLC.



Published online: 09 Oct 2023.



Submit your article to this journal [↗](#)



Article views: 107



View related articles [↗](#)



View Crossmark data [↗](#)

discussion that Porsdam Mann and colleagues have initiated.

## FUNDING

The author reported there is no funding associated with the work featured in this article.

## ORCID

Alexandre Erler  <http://orcid.org/0000-0001-5553-7567>

## REFERENCES

- Checco, A., L. Bracciale, P. Loreti, S. Pinfield, and G. Bianchi. 2021. AI-assisted Peer review. *Humanities & Social Sciences Communications* 8. doi:10.1057/s41599-020-00703-8.
- Flaherty, C. 2022. The peer-review crisis. *Inside Higher Ed* [Online]. Available: <https://www.insidehighered.com/news/2022/06/13/peer-review-crisis-creates-problems-journals-and-scholars> [Accessed 06/08/2023].
- McLean, S., G. J. M. Read, J. Thompson, C. Baber, N. A. Stanton, and P. M. Salmon. 2023. The risks associated with artificial general intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence* 35 (5):649–63. doi:10.1080/0952813X.2021.1964003.
- Nam, J., S. Mo, J. Lee, and J. Shin. 2023. Breaking the spurious causality of conditional generation via fairness intervention with corrective sampling. arXiv:2212.02090 [cs.CV]. doi:10.48550/arXiv.2212.02090.
- Porsdam Mann, S., B. D. Earp, N. Möller, S. Vynn, and J. Savulescu. 2023. AUTOGEN: A personalized large language model for academic enhancement—ethics and proof of principle. *The American Journal of Bioethics* 23 (10):28–41. doi:10.1080/15265161.2023.2233356.
- Weber-Wulff, D., A. Anohina-Naumeca, S. Bjelobaba, T. Foltýnek, J. Guerrero-Dib, O. Popoola, P. Šigut, & L. Wadding. 2023. Testing of detection tools for AI-generated text. arXiv:2306.15666 [cs.CL]. doi:10.48550/arXiv.2306.15666.
- Zupanc, G. K. H. 2023. It is becoming increasingly difficult to find reviewers—Myths and facts about Peer review. *Journal of Comparative Physiology A*. doi:10.1007/s00359-023-01642-w.

THE AMERICAN JOURNAL OF BIOETHICS  
2023, VOL. 23, NO. 10, 96–99  
<https://doi.org/10.1080/15265161.2023.2250315>




Taylor & Francis  
Taylor & Francis Group

## OPEN PEER COMMENTARIES



# Reimagining Scholarship: A Response to the Ethical Concerns of AUTOGEN

Hazem Zohny 

Oxford Uehiro Centre for Practical Ethics, University of Oxford

In their recent paper “AUTOGEN: A Personalized Large Language Model for Academic Enhancement—Ethics and Proof of Principle,” Porsdam Mann et al. (2023) demonstrate a technique for fine-tuning the large language model (LLM) GPT-3, allowing it to generate academic text that appears personalized to an author’s unique scholarly voice.

LLMs introduce numerous considerations for academic research. On one hand, they promise substantial improvements in academic writing productivity and efficiency. They could save considerable time in drafting and editing, broaden authorship opportunities for non-native English speakers, and function as a Devil’s Advocate by instantly producing

counterarguments during the drafting process (Zohny, McMillan, and King 2023).

However, users may also fine-tune LLMs on prolific, established authors’ writings rather than on their own, or they may fine-tune them on a blend of top-tier publications, aiming to optimize academic content generation for specific topics. These would be closer to “impersonator” or “bespoke” rather than personalized LLMs.

These possibilities may seem to flare up the ethical concerns Porsdam Mann and colleagues raise, but I will argue at least some of these can be addressed. In what follows, I challenge some of their worries specifically around the consensual use of publications, the

CONTACT Hazem Zohny  [hazem.zohny@philosophy.ox.ac.uk](mailto:hazem.zohny@philosophy.ox.ac.uk)  Oxford Uehiro Centre for Practical Ethics, University of Oxford, Oxford, UK.

© 2023 Taylor & Francis Group, LLC.

“Matthew effect,” plagiarism, homogenization, and instrumentalization and alienation.

## CONSENT

The authors raise a number of concerns around consent in the context of fine-tuned LLMs, especially related to the use of writings and writing styles without permission.

It’s worth firstly noting that “writing styles” and (to the extent that we can identify them) “thinking styles,” are not subject to copyright as these are methods or techniques for expressing an idea (US Copyright Office [n.d.](#)). There is no requirement to seek permission before, for instance, publishing a philosophical work that mimics the writing style of, say, Derek Parfit’s distinctive blend of precision, rigorous logical structuring, and use of reductive analysis.

How copyright law will evolve in response to generative AI remains to be seen (Samuelson 2023), but while copyright safeguards creative expression, the process of model training focuses on extracting non-protected concepts and trends from data (Quang 2023). Using the academic texts of others to reshuffle and transmute them via LLMs could be seen as “transformative work.” Transformative work allows the unlicensed use of copyrighted material for “criticism, comment, news reporting, teaching, scholarship, and research” under fair use (Copyright Act, 17 U.S.C. § 107).

From a consent perspective, this can be analogized to authors who incorporate various influences into their writings without seeking explicit permission. Similarly, LLMs can blend a plethora of texts to produce distinct outputs. Even if an LLM might replicate sections of copyrighted works word-for-word, the remedy would be akin to current practices: ensuring proper citation of original authors via peer review and plagiarism detection software.

In contrast, if a bad actor used an LLM to merely rephrase a publication while attributing it to themselves, that would be closer to a derivative work that may violate copyright protection, and would be considered plagiarism (more on that below).

## THE MATTHEW EFFECT

The authors worry that established academics, who already enjoy the benefits of longer publication records, can benefit disproportionately from personalized LLMs. They may be able to further cement their

advantage by leveraging personalized LLMs to be even more productive in generating new manuscripts.

However, the possibility of fine-tuning LLMs on the writings of other authors could largely neutralize this issue. Training impersonator or bespoke models on prolific academics’ writings may actually help level the playing field. If early career researchers with limited publications can utilize top scholars’ publicly available publications, this could close capability gaps that would otherwise advantage established academics. Doing so need not preclude the same authors from also using an LLM personalized to their own writing style.

## HOMOGENIZATION

Another ethical concern the authors raise is that personalized LLMs could lead to the homogenization of writing styles, potentially inhibiting individual stylistic evolution, reducing collective diversity, and diminishing the authenticity and enjoyment of LLM-produced texts. This possibility seems relevant to nonacademic writing where authenticity and the esthetic value of a writing style are more important.

The goal of academic writing, in contrast, is to disseminate knowledge effectively. Academic writing is notorious for being esoteric and cumbersome (Steven 2014). If LLMs have the effect of homogenizing academic writing styles while in the process making them more accessible and coherent, this may be a worthwhile tradeoff. Especially for disciplines like applied ethics, where the goal is to achieve conceptual clarity and illuminate value tradeoffs, a uniform, clear, and direct writing style would be a welcome development.

On the other hand, this concern might become less relevant when we consider the likely role LLMs will play in the consumption (as opposed to production) of academic work. Readers may increasingly leverage LLMs to digest academic papers in their preferred formats and styles—they might, for instance, request LLMs to transmute articles or passages into stylized dialogues or vignettes, effectively customizing their reading experience. This is already possible, with LLMs such as Anthropic’s Claude.ai allowing users to attach several large PDFs at a time, and then creating detailed debates between the authors of each paper. The power of customization here can offer each reader a unique experience of an argument, effectively diversifying rather than homogenizing academic content.

## PLAGIARISM

As Porsdam Mann et al. (2023) note, there are likely to be new challenges over potential plagiarism and authorship misrepresentation through unauthorized imitation of writing styles facilitated by these LLMs.

It's worth firstly noting that writing styles cannot be plagiarized in the traditional sense of the term, which involves presenting someone else's specific content as one's own. Moreover, there are already established mechanisms to counteract instances where fine-tuned models might reproduce passages verbatim from another author without appropriate citations. For instance, academic journals routinely utilize software like CrossCheck and iThenticate (Carter and Blanford 2016).

A more pressing issue is the ease with which LLMs enable semantic plagiarism, where ideas are rephrased without appropriate citation. Yet, the very technology causing these concerns might also offer solutions. Take for example Jenni.ai, a generative AI writing assistant. Users can highlight passages in their manuscript, and the tool suggests relevant journal articles from its linked database. As LLMs advance and are fine-tuned for various literatures (Gao et al. 2023), they could automatically cite sources when recognizing ideas found in previous works. Such mechanisms could be seamlessly integrated into both writing and peer review workflows.

## INSTRUMENTALIZATION & ALIENATION

The authors briefly allude to the risk that LLMs could instrumentalize and alienate academic writers but do not delve deeply into these issues. They hint that if LLMs can generate academic texts and ideas with minimal effort, this could inadvertently transform scholars into mere operators of these models, potentially creating a rift between them and the creative academic process.

Current LLMs cannot produce publishable academic writing without substantial human help, so instrumentalization concerns may be premature now. Still, it's prudent to consider future implications if LLMs grow more powerful. Scholars and LLMs could form a pilot-copilot relationship with scholars guiding the trajectories of their disciplines at a high level.

However, another possibility is that scholars will eventually be reduced to the equivalent of assembly line workers—they merely monitor the prompts that are directing LLMs, where those prompts themselves have been fed to them by other LLMs, and they

supervise outputs that are part of a larger much work that they do not understand. That might sound like a dystopian development considering the historical connotations of assembly line work with mindless toil and exploitation. However, reframing this comparison with how assembly lines can enhance productivity and quality may yield a compelling argument in favor of the diminished shift in the role of scholars.

This is because assembly line work isn't merely about fragmentation of tasks and alienation of workers, but it is also a well-orchestrated system that allows for efficiency, consistency, and high-quality production. Similarly, the transformation of scholars into instrumentalized operators of LLMs, while potentially stripping them of their creative role, could lead to an explosion in the productivity, reach, and impact of research. Reimagining scholarship in this way sounds deeply unappealing, but it may be we are heading toward a tradeoff situation between accelerated knowledge discovery and dissemination on one hand, and the preservation of interesting, rewarding scholarly jobs on the other.

No doubt, any such development would raise its own risks: pre-trained flaws propagating through disciplines, excessive convergence of ideas, loss of risky, heterodox or speculative research, or other unforeseen consequences. The point here is not to dismiss these possibilities but to highlight that scholars finding their work less creative or rewarding due to LLMs may signal that these systems are progressing scholarly fields better than their human counterparts. Of course, transitioning to LLMs just to cut costs *and* at the expense of quality and rigor of academic work would indeed be a poor tradeoff.

In sum, while LLMs and their fine-tuning raise concerns for academic writing and progress, at least some of these can be addressed through the same ongoing developments in AI. Continued interdisciplinary dialogue on the sociotechnical implications of LLMs will be vital to their responsible integration into academia.

## DISCLOSURE STATEMENT

No potential conflict of interest was reported by the author(s).

## FUNDING

No funding was received in association with the work featured in this article.

## ORCID

Hazem Zohny  <http://orcid.org/0000-0002-7734-2186>

## REFERENCES

- '17 U.S. Code § 107 - Limitations on Exclusive Rights: Fair Use'. n.d. LII/Legal Information Institute. Accessed 6 August 2023. <https://www.law.cornell.edu/uscode/text/17/107>.
- Carter, C. B., and C. F. Blanford. 2016. Plagiarism and detection. *Journal of Materials Science* 51 (15):7047–8. doi:10.1007/s10853-016-0004-7.
- Gao, T., H. Yen, J. Yu, and D. Chen. 2023. Enabling large language models to generate text with citations. *arXiv*. doi:10.48550/arXiv.2305.14627.
- Porsdam Mann, S., B. Earp, N. Moller, V. Suren, and J. Savulescu. 2023. AUTOGEN: A personalized large language model for academic enhancement—Ethics and proof of principle. *American Journal of Bioethics* 23 (10): 28–41. doi:10.1080/15265161.2023.2233356.
- Quang, J. 2023. Does training AI violate copyright law? *Berkeley Technology Law Journal* 36 (4):1407–36.
- Samuelson, P. 2023. Generative AI meets copyright. *Science* (New York, N.Y.) 381 (6654):158–61. doi:10.1126/science.adi0656.
- Steven, P. 2014. 'Why Academics Stink at Writing'. *The Chronicle of Higher Education*, 2014, sec. The Review. <https://www.chronicle.com/article/why-academics-stink-at-writing/>.
- US Copyright Office. n.d. 'What does copyright protect? (FAQ) | U.S. copyright office'. Web page. Accessed 6 August 2023. <https://www.copyright.gov/help/faq/faq-protection.html>.
- Zohny, H., J. McMillan, and M. King. 2023. Ethics of generative AI. *Journal of Medical Ethics* 49 (2):79–80. doi:10.1136/jme-2023-108909.

THE AMERICAN JOURNAL OF BIOETHICS  
2023, VOL. 23, NO. 10, 99–102  
<https://doi.org/10.1080/15265161.2023.2250319>




Taylor & Francis  
Taylor & Francis Group

## OPEN PEER COMMENTARIES



## Generative AI and the Foregrounding of Epistemic Injustice in Bioethics

Calvin Wai-Loon Ho 

University of Hong Kong

OpenAI's Chat Generative Pre-training Transformer (ChatGPT), Google's Bard and other generative artificial intelligence (GenAI) technologies can greatly enhance the capability of healthcare professionals to interpret data across different data sources and locations with a simple query, as well as advance medical research through its ability to generate synthetic data (The Lancet Regional Health-Europe 2023). However, the performance of these technologies depends on the data they are trained on. Existing data may be seriously biased due to a lack of gender, ethnic, racial, social and/or religious diversity, and is a concern that the Global Alliance for Genomics & Health (2023) seeks to address in a recent initiative to promote global diversity in datasets within genomic research. If used in clinical medicine, the results from GenAI technologies present serious normative challenges that Cohen (2023) has clearly and succinctly set out, quite

aside from the direct impact that they could have on human health and wellbeing.

While it should come as no surprise to anyone that emerging health technologies tend to present normative and regulatory challenges, many of the "new-ish" problems that are anticipated to arise from the use of GenAI technologies in healthcare and research foreground intransigent concerns with epistemic injustice. I provide three reasons why GenAI's clinical use is a big deal in bioethics. First, it highlights that bioethics does not adequately account for the impact that power dynamics and systemic biases have in knowledge production and dissemination. Marginalized individuals and communities still lack the capability to participate in knowledge production and decision-making processes, even if well-formulated informed consent procedures are in place. Second, there may be a disconnect between bioethics and the lived experiences of

**CONTACT** Calvin Wai-Loon Ho  [cwlho@hku.hk](mailto:cwlho@hku.hk)  Faculty of Law, University of Hong Kong, Hong Kong, China.

© 2023 Taylor & Francis Group, LLC