Transparent disclosure of artificial intelligence tools in writing scholarly manuscripts

Mohammad Hosseini, Kristi Holmes

Background

Since November 2022 when ChatGPT was released, the academic community has had to deal with novel ethical issues in education and research. The use of artificial intelligence tools and systems based on large language models (LLMs) in scholarly outputs has caught journal editors, publishers and policy developers off-guard. Some journals (e.g., Science) and events (e.g., International Conference on Machine Learning) banned the use of LLMs but others developed suggestions on how to use LLMs in a responsible manner.
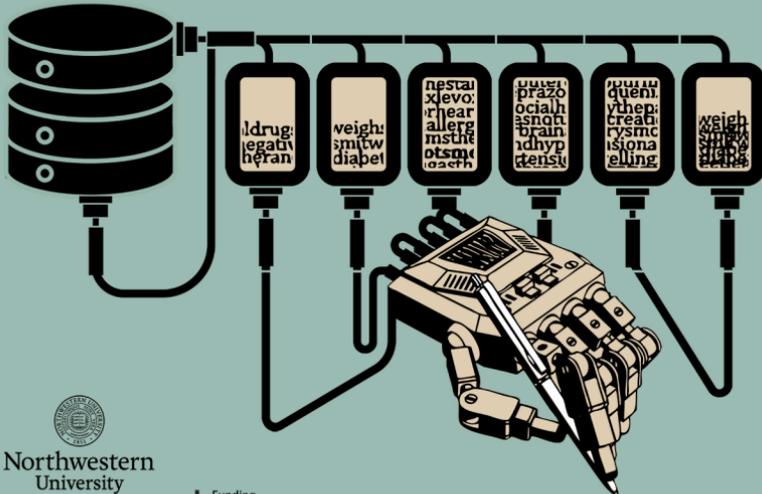
Objectives and methods

We conceptually explored the ethics of disclosing the use of LLMs in research to develop best practices for researchers in all scholarly fields. We identify four different approaches towards disclosure of LLMs' use in scholarly publications and highlight specific problems associated with them (table 1).

Table 1. Evaluation of different policy options concerning the use AI in writing or editing scholarly publications.

Policy Option	Rationale	Problems
Ban the use of AI in generating texts for scholarly manuscripts	> Avoids difficult issues related to fair allocation of authorship credit, accountability, and transparency	> Not enforceable > Leads to clandestine use of AIs > Discourages equity in science and prevents helping researchers who are not adept at writing in languages other than their first language
Allow AIs to be listed as authors	> Avoids giving human authors undue credit for work done by AIs > Promotes transparency	> AIs cannot be morally or legally responsible or accountable
Allow AIs to be listed in the acknowledgments section	> Promotes transparency	> AIs cannot be morally or legally responsible or accountable
Disclose use of AIs in the body of the texts and among references	> Promotes transparency > Consistent with disclosing the use of other tools	> Consistency of disclosure

Results

LLMs should not be named as authors nor mentioned in the acknowledgments because they do not have free will and therefore cannot be held morally or legally responsible for what they do. Use of LLMs, like the use of other software tools, should be cited in-text, followed by being mentioned in the references.

Best practices for disclosure of using LLMs

To uphold ethical norms of transparency, openness, honesty, and fair attribution of credit, in cases where LLMs are used, disclosure should happen:

1) As free text in the introduction or methods section (to honestly and transparently describe details about who used LLMs, when, how, using what prompts and disclose what sections of the text are affected; to prevent giving undue credit to human contributors for work they did not do)

2) Through in-text citations and among references (to improve findability and indexing) using the following format: OpenAI (2023). ChatGPT (GPT-4, May 12 Version) [Large language model]. Response to query made by X.Y. Month/Day/Year. https://chat.openai.com/chat

3) To enable verification, interactions with LLMs (including specific prompts, and dates of query) should be recorded and disclosed as supplementary material or in appendices.

Conclusions

We believe that bans are unenforceable and may encourage undisclosed use of LLMs. Furthermore, in experimenting with different applications based on LLMs, we have witnessed their remarkable capabilities in assisting researchers with writing, reviewing, and editing text. For example, LLMs can be particularly helpful in improving science communication and tasks such as translation and summarization. Accordingly, these tools can level the playing field for those engaged in research and education in a language other than their first language and promote equity in science.

Northwestern University

NUCATS Clinical and Translational Sciences Institute