Using ChatGPT to conduct a literature review

Michael Haman & Milan Školník

Taylor & Francis
Taylor & Francis Group

LETTER TO THE EDITOR

Using ChatGPT to conduct a literature review

Michael Haman ⓘ and Milan Školník ⓘ

Department of Humanities, Czech University of Life Sciences Prague; Prague, Czech Republic

Dear Accountability in Research Editors,

In this letter, we consider the ethical and practical implications of using ChatGPT in the research process. The ethics of using ChatGPT have been debated in both the public and academic literature since its inception (Hosseini, Rasmussen, and Resnik 2023). However, only some dare to suggest that tools similar to ChatGPT could replace the entire research process. Some views have been expressed that similar services could replace at least part of the research process. In this letter, we focus on how ChatGPT can currently perform a literature review. The first thing to consider is the ethical side of the matter in general. Is it ethical to use AI services to conduct a literature review? One could argue that this is a fundamental process of research activity; if ChatGPT cannot coauthor, then it is unethical to use ChatGPT to produce a literature review. The authors of a publication must be held responsible for producing the literature review, but AI can in no way be held responsible for this activity.

Apart from the ethical aspect, it is also necessary to consider whether using ChatGPT for a literature review is practical. We used ChatGPT on 23 February 2023. ChatGPT was given the prompt (Figure 1): "'List 10 seminal academic articles in the field of medicine and provide the DOI.'" We used this prompt five times, resulting in five unique chats. ChatGPT provided us with 10 publications each time, for 48 unique publications from the five attempts. We found that only 8 of the 50 provided DOIs exist and are listed for the correct publication. At the same time, we searched for the title of the paper itself with the authors and the correct academic publication (even though the DOI may have been wrong), and only in 17 out of 50 cases did the article in question exist in the databases (Google Scholar, PubMed, Semantic Scholar). In 66% of the cases, ChatGPT gave us a non-existent paper. Considering that the prompt was general to the whole field of medicine and not a prompt on specific research questions, these are very poor results. ChatGPT might have given much worse results if we had wanted to focus on a specific research question.

The short tests indicate that ChatGPT does not stand up to a literature review because it gives too many fake papers. The number of fake papers makes it difficult to even recommend ChatGPT as an aid

M   List 10 seminal academic articles in the field of medicine and provide the DOI

Sure, here are 10 seminal academic articles in the field of medicine, along with their DOI:

1. Charles R. Drew, "Banked Blood," Journal of the National Medical Association, vol. 28, no. 4 (1936): 123-126. DOI: 10.1007/s00464-020-07623-8
2. Jonas Salk, "The Challenge of Poliomyelitis," Journal of the American Medical Association, vol. 159, no. 4 (1955): 318-324. DOI: 10.1001/jama.1955.02950220002001
3. Roger Guillemin, Andrew V. Schally, and Rosalyn S. Yalow, "Isolation, structure, and synthesis of pituitary and hypothalamic hormones," Journal of Biological Chemistry, vol. 242, no. 22 (1967): 5747-5752. DOI: 10.1016/0006-291x(68)90207-5
4. David A. Kessler, "Introducing MEDLINEPlus," Bulletin of the Medical Library Association, vol. 86, no. 3 (1998): 427-428. DOI: 10.5860/choice.37-0806
5. Michael S. Lauer and Harlan M. Krumholz, "Top 10 Myths Regarding the Diagnosis and Treatment of Myocardial Infarction," Journal of the American Medical Association, vol. 286, no. 7 (2001): 988-992. DOI: 10.1001/jama.286.7.988
6. Stephen B. Soumerai and Dennis Ross-Degnan, "Advances in Pharmacotherapy for Hypertension: Implications for Health Policy," Health Affairs, vol. 23, no. 3 (2004): 24-40. DOI: 10.1377/hlthaff.23.3.24
7. Edward H. Livingston and Gregory L. Fricchione, "Ethics and the Pharmaceutical Industry," New England Journal of Medicine, vol. 355, no. 14 (2006): 1499-1502. DOI: 10.1056/nejmp068182
8. Frank G. Holz, Daniel Pauleikhoff, and Alan C. Bird, "Pharmacological Therapy of Age-Related Macular Degeneration," Ophthalmology, vol. 113, no. 5 (2006): 874-882. DOI: 10.1016/j.ophtha.2006.01.067
9. Eric J. Topol, "Individualized Medicine from Pre-Womb to Tomb," Cell, vol. 157, no. 1 (2014): 241-253. DOI: 10.1016/j.cell.2014.02.012
10. Thomas J. Moore and Curt D. Furberg, "Development and Regulation of Modern Antidepressants," Archives of General Psychiatry, vol. 63, no. 7 (2006): 716-724. DOI: 10.1001/archpsyc.63.7.716

Figure 1. A prompt used in the test.

in the research process. In short, nothing that ChatGPT sovereignly writes can be trusted. Sovereignty is a big problem. If ChatGPT delivered the existing papers and wrote about the rest saying that it does not know about other papers, it would not be such a problem. At the same time, it is also able to write about each publication and what such publications are about. Therefore, we firmly recommend not using ChatGPT in the research process.

Disclosure statement

Funding

ORCID

Michael Haman http://orcid.org/0000-0001-5772-2045
Milan Školník http://orcid.org/0000-0002-0672-219X

Reference

Hosseini, M., L. M. Rasmussen, and D. B. Resnik. 2023. "Using AI to Write Scholarly Publications." *Accountability in Research* 1–9. doi:10.1080/08989621.2023.2168535.