

Letter to the Editor

Citation Accuracy Challenges Posed by Large Language Models

Manlin Zhang^{1*}, MPhil; Tianyu Zhao^{2*}, MSc

¹Department of Obstetrics and Gynecology, Shengjing Hospital of China Medical University, Shenyang, China

²Department of Science and Technology Studies, University College London, London, United Kingdom

*all authors contributed equally

Corresponding Author:

Tianyu Zhao, MSc
Department of Science and Technology Studies
University College London
Gower Street
London, WC1E 6BT
United Kingdom
Phone: 86 13674280942
Email: tianyu.zhao.23@ucl.ac.uk

Related Articles:

Comment on: <https://mededu.jmir.org/2025/1/e63400>

Comment in: <https://mededu.jmir.org/2025/1/e73698>

JMIR Med Educ 2025;11:e72998; doi: [10.2196/72998](https://doi.org/10.2196/72998)

Keywords: chatGPT; medical education; Saudi Arabia; perceptions; knowledge; medical students; faculty; chatbot; qualitative study; artificial intelligence; AI; AI-based tools; universities; thematic analysis; learning; satisfaction; LLM; large language model

Large language models (LLMs) such as DeepSeek, ChatGPT, and ChatGLM have significant limitations in generating citations, raising concerns about the quality and reliability of academic research. These models tend to produce citations that are correctly formatted but fictional in content, misleading users and undermining academic rigor. In the recent study titled “Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: qualitative study,” the section addressing concerns about ChatGPT deserves a deeper discussion [1].

There are several reasons for the citation issues in LLMs, which can be analyzed as follows. First, most LLMs cannot access paid subscription databases and therefore solely rely on open-access resources [2]. This limits the citations generated by LLMs to open-access journals, potentially omitting more significant research published in subscription-based journals. Second, LLMs are trained on vast amounts of text data and generate content by analyzing patterns and structures in text. However, they lack the ability to understand the content of the text or think critically, implying that they cannot judge the accuracy and reliability of information. Third, the algorithms underlying LLMs are often opaque, leaving users unable to understand the specific processes of information handling. This makes it difficult for users

to determine the reliability of citations generated by LLMs and to effectively evaluate their results. Recent research also stated that half of generated search results lack citations, and only 75% of those with citations support the claims, posing trust concerns as user reliance grows[3].

Recently, an experiment conducted by the Journal of Clinical Anesthesia involved publishing a fictional article titled “Spinal Cord Ischemia After ESP Block” to test the spread and citation of a fabricated academic content. Surprisingly, the fictional article was widely cited, over 400 times, including in some journals with high impact factors[4], revealing a lack of rigor in academic citation practices, where many authors may not check the original literature and instead copy references directly. This incident sparked widespread discussion about academic citation practices, emphasizing the importance of critical thinking by scholars while citing materials.

The use of fictional citations by LLMs poses a multi-faceted problem: it misleads users into drawing incorrect conclusions and making inappropriate decisions, undermines the rigor and credibility of academic research, and hinders the dissemination of knowledge by limiting access to accurate scientific information [5]. The issue of LLMs generating fictional citations is complex and requires the

combined efforts of multiple stakeholders for resolution. Developers must continuously improve the LLM technology and algorithms, users must increase their awareness and critical evaluation skills while using LLMs, and academic

institutions must strengthen the management and education in academic practices. Only through these efforts can we ensure that LLMs play a positive role in academic research and promote the dissemination and progress of knowledge.

Conflicts of Interest

None declared.

References

1. Abouammoh N, Alhasan K, Aljamaan F, et al. Perceptions and earliest experiences of medical students and faculty with ChatGPT in medical education: qualitative study. *JMIR Med Educ.* Feb 20, 2025;11:e63400. [doi: [10.2196/63400](https://doi.org/10.2196/63400)] [Medline: [39977012](https://pubmed.ncbi.nlm.nih.gov/39977012/)]
2. Perianes-Rodríguez A, Olmeda-Gómez C. Effects of journal choice on the visibility of scientific publications: a comparison between subscription-based and full open access models. *Scientometrics.* Dec 2019;121(3):1737-1752. [doi: [10.1007/s11192-019-03265-y](https://doi.org/10.1007/s11192-019-03265-y)]
3. Peskoff D, Stewart B. Credible without credit: domain experts assess generative language models. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics.* Jul 2023;2:427-438. [doi: [10.18653/v1/2023.acl-short.37](https://doi.org/10.18653/v1/2023.acl-short.37)]
4. Marcus A, Oransky I, De Cassai A. Please don't cite this editorial. *J Clin Anesth.* Jan 8, 2025;111741. [doi: [10.1016/j.jclinane.2025.111741](https://doi.org/10.1016/j.jclinane.2025.111741)] [Medline: [39779384](https://pubmed.ncbi.nlm.nih.gov/39779384/)]
5. Rasul T, Nair S, Kalendra D, et al. The role of ChatGPT in higher education: benefits, challenges, and future research directions. *JALT.* May 10, 2023;6(1):41-56. [doi: [10.37074/jalt.2023.6.1.29](https://doi.org/10.37074/jalt.2023.6.1.29)]

Abbreviations

LLM: large language model

Edited by Surya Nedunchezhiyan; This is a non-peer-reviewed article; submitted 23.02.2025; accepted 12.03.2025; published 02.04.2025

Please cite as:

Zhang M, Zhao T

Citation Accuracy Challenges Posed by Large Language Models

JMIR Med Educ 2025;11:e72998

URL: <https://mededu.jmir.org/2025/1/e72998>

doi: [10.2196/72998](https://doi.org/10.2196/72998)

©Manlin Zhang, Tianyu Zhao. Originally published in *JMIR Medical Education* (<https://mededu.jmir.org>), 02.04.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in *JMIR Medical Education*, is properly cited. The complete bibliographic information, a link to the original publication on <https://mededu.jmir.org/>, as well as this copyright and license information must be included.