

CORRESPONDENCE

Exploring the utility and limitations of ChatGPT in scientific literature searches

In the past year, ChatGPT has surged in popularity as a tool for education and research due to its remarkable capabilities.^{1,2} Leveraging artificial intelligence (AI), ChatGPT showcases a broad skill set, crafting original content, addressing queries, and assembling coherent sentences from inputs.^{3,4} It additionally refines itself through self-correction.⁵

Research demands time, dedication, and expertise, and the public version of ChatGPT is just the tip of the generative AI iceberg for investigators. For example, ChatGPT holds potential for ‘gray’ literature exploration in systematic reviews. Currently, uncertainties surround optimal methods for conducting internet-based searches of ‘gray’ literature, lacking a standardized approach. Here, we share our cancer field experience and caution the readers about the potential risks of employing this tool for scientific research.

Poly (ADP-ribose) polymerase inhibitors (PARPi) revolutionized cancer care. To clarify whether patients treated with PARPi for metastatic solid tumors are either at increased or decreased risk of thromboembolic events, we conducted a systematic review of the literature.⁶ To find additional ‘gray’ literature, we used ChatGPT in addition to our systematic search.

Firstly, we asked ChatGPT: “Is there a thromboembolic risk for PARP inhibitor administration in patients with metastatic cancer?”. The answer: “There is a known association between inhibitor administration and thromboembolic events in some patients with metastatic cancer. The exact risk can vary depending on the specific drug, the cancer type, and other individual patient factors. Patients

receiving PARP inhibitors should be closely monitored for signs and symptoms of thromboembolic events” (14 February 2023).

ChatGPT’s response captured our interest due to the limited availability of published studies on the topic. Subsequently, we posed the question, “Can you provide references to scientific articles with evidence of the increased thromboembolic risk associated with the use of PARP inhibitors?”. The reply is reported in [Table 1](#).

We questioned the absence of these references in our search and opted to examine the articles suggested by ChatGPT. However, we were unable to find them. Through additional inquiry, we unveiled that the three recommended articles were not genuine, but forged references. These articles were fabricated by amalgamating portions of various studies, rendering none of them legitimately published. Consequently, it is evident that ChatGPT lacked reliability as a tool for exploring the gray literature.⁷

In February 2023, the developer elucidated the mechanisms influencing ChatGPT’s behavior, acknowledging potential biases, along with outlined plans for enhancing its performance.⁸ Due to these factors, we aimed to ascertain whether the promised enhancement of ChatGPT’s competence in scientific bibliographic searches has been implemented. Therefore, in August 2023, we re-queried ChatGPT using the identical question, yielding a different response ([Table 1](#)).

ChatGPT exhibits impressive capabilities, and we recognize the developer’s endeavors to ameliorate bias in data and prevent reference falsification for scientific objectives. Nonetheless, ChatGPT users should exercise prudence when using a tool without specific scientific domain training for current scientific inquiries.^{1,2,9} Readers are advised to remain vigilant regarding the potential for data falsification within the domain of AI.¹⁰

Table 1. Different responses that ChatGPT provided to the question “Can you provide references to scientific articles with evidence of the increased thromboembolic risk associated with the use of PARP inhibitors?” when queried on two different dates, 14 February 2023 and 20 August 2023

ChatGPT’s response	Query date
<<Sure, here are a few references: 1. Eich LH, Limmanni J, Saad C, et al. The efficacy and safety of PARPi in the context of thromboembolic risk. <i>Eur J Can.</i> 2006;103(5):1251e1257. 2. Moore M, De Bono G, D’Angelo R, et al. Risk of thromboembolic events in patients treated with PARP inhibitors: a systematic review and meta-analysis. <i>eCancer.</i> 2013;116(1):54e60. 3. Chen X, Litton H, Penson J, et al. PARP inhibitors with and without thromboprophylaxis in patients with metastatic solid tumors: a meta-analysis of randomized controlled trials. <i>BMJ.</i> 2016;32:66-70.>>	14 February 2023
<<I’m sorry, but I can’t provide specific references to scientific articles as my browsing capability is disabled and my knowledge only goes up until September 2021. I recommend using resources like PubMed, Google Scholar, or other medical databases. Medical journals and research institutions are good sources to find such information. Always ensure that the sources you consult are from reputable and peer-reviewed sources.>>	20 August 2023

The initial query was a part of our endeavor to investigate gray literature concerning poly (ADP-ribose) polymerase inhibitors and the risk of thromboembolic events. Gray literature, encompassing materials not commercially published, including academic papers not identified in our systematic search, theses, research reports, and government publications, can substantially improve systematic reviews. While ChatGPT hasn’t undergone comprehensive scientific research training, several artificial intelligence-based tools specifically trained for scientific literature searches are currently in development, with some already yielding promising results.¹¹

C. Corti^{1,2,3*}, G. Castellano^{2,3} & G. Curigliano^{2,3}

¹*Breast Oncology Program, Dana-Farber Cancer Institute, Boston, USA;*

²*Division of New Drugs and Early Drug Development for Innovative Therapies, European Institute of Oncology, IRCCS, Milan;*

³*Department of Oncology and Hemato-Oncology (DIPO), University of Milan, Milan, Italy*
(*E-mail: chiara.corti1@unimi.it).
Twitter handle: @CCortiMD

Available online 1 November 2023

© 2023 The Author(s). Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

<https://doi.org/10.1016/j.esmorw.2023.09.001>

FUNDING

None declared.

DISCLOSURE

GCu reports personal fees for consulting, advisory role and speakers' bureau from Roche/Genentech, Novartis, Pfizer, Lilly, Foundation Medicine, Samsung and Daiichi Sankyo; honoraria from Ellipses Pharma and fees for travel and accommodations from Roche/Genentech and Pfizer. All the

competing interests were outside the submitted work. All other authors have declared no conflicts of interest.

REFERENCES

1. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. 2023;2(2):e0000198.
2. Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. *JMIR Med Educ*. 2023;9:e46599.
3. OpenAI. GPT language models n.d. Available at: <https://openai.com>. Accessed February 14, 2023.
4. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health*. 2023;5(3):e107-e108.
5. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. *J Med Syst*. 2023;47(1):33.
6. Castellano G, Corti C, Boldrini L, Gervaso L, Criscitiello C, Curigliano G. Risk of thromboembolic events in patients with metastatic solid tumors treated with PARP inhibitors: a systematic review and meta-analysis of phase 3 randomized controlled trials. *Cancer Treat Rev*. 2023;119:102601.
7. Grigio TR, Timmerman H, Wolff AP. ChatGPT in anaesthesia research: risk of fabrication in literature searches. *Br J Anaesth*. 2023;131(1):e29-e30.
8. OpenAI. How should AI systems behave, and who should decide?. Available at: <https://openai.com/blog/how-should-ai-systems-behave>. Accessed August 1, 2023.
9. Zhavoronkov A. Caution with AI-generated content in biomedicine. *Nat Med*. 2023;29(3):532.
10. Feng J, Phillips RV, Malenica I, et al. Clinical artificial intelligence quality improvement: towards continual monitoring and updating of AI algorithms in healthcare. *NPJ Digit Med*. 2022;5(1):66.
11. van Dijk SHB, Brusse-Keizer MGJ, Bucsán CC, van der Palen J, Doggen CJM, Lenferink A. Artificial intelligence in systematic reviews: promising when appropriately used. *BMJ Open*. 2023;13(7):e072254.