

Use of large language models for evidence-based cardiovascular medicine

Ioannis Skalidis , Aurelien Cagnina, and Stephane Fournier *

Department of Cardiology, University Hospital of Lausanne, Rue du Bugnon 46, 1011 Lausanne, Switzerland

Online publish-ahead-of-print 17 July 2023

Commentary article to: ‘ChatGPT fails the test of evidence-based medicine’, by W. Haverkamp et al. <https://doi.org/10.1093/ehjdh/ztad043>.

We would like to thank Haverkamp et al.¹ for their interesting comment on our commentary article titled ‘ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story?’.² While the current capabilities of Chat Generative Pre-trained Transformer (GPT) do not meet the stringent requirements to deliver unsupervised evidence-based medicine recommendations, we believe that this technology can nevertheless help the physicians to reach evidence-based medicine. With further advancements and rigorous evaluation, it has the potential to contribute and find a role in clinical practice in the future.

We concur with the authors that the euphoric tone surrounding ChatGPT’s performance in passing medical exams should be balanced with critical analysis^{3,4} and would like to emphasize that fully aware of important current limitations, our commentary article was published with a question mark in its title. We also recently published a viewpoint article illustrating the limitations of ChatGPT by showing how small changes in the way questions are asked can result in significantly different outputs and raised some questions about the legal consequences associated with the use of these artificial intelligence (AI) tools.⁵

While it provided correct answers to 60% of the sampled questions of the European Exam in Core Cardiology, the remaining 40% of incorrect or indeterminate answers cannot be overlooked.^{4,6} We fully agree that even a small percentage of incorrect answers can pose significant risks if the system is relied upon excessively. Moreover, we should

not forget that crucial and considerably more demanding questions including videos (coronary angiography, echocardiography, magnetic resonance imaging, etc.) and pictures (electrocardiograms, pacemaker tracings, etc.) were excluded.

We share the concern regarding the lack of transparency surrounding ChatGPT’s training data and methods. The absence of extensive documentation and undisclosed sources limit our ability to evaluate the system’s reliability accurately.⁷ Additionally, the authors rightly emphasize that evidence-based medicine necessitates transparent incorporation of medical guidelines, selection of sources, and up-to-date information. ChatGPT’s inability to specify the extent to which medical guidelines have been incorporated is a valid concern, highlighting the need for transparency and comprehensibility in its training process.

However, we certainly do not believe that the role of ChatGPT is to replace doctors, but we believe it can assist doctors in their daily lives. It is easy to imagine a potential benefit in asking the AI for a differential diagnosis of a clinical situation, for example. In this situation, the doctors could bring a critical evaluation to the suggested list, which may propose ideas they had not thought of. We can also imagine that submitting a list of medications to ChatGPT in a particular scenario could potentially highlight an interaction that had not been seen, or suggest a medication that could be useful.⁸ The prescriber will obviously always have the final say, but the AI can serve as a safety net and in this sense likely contribute to delivering better evidence-based medicine (Figure 1).⁹

Is ChatGPT useful in assisting physicians to deliver the best possible evidence-based medical treatment?

Yes.



Figure 1 A question addressed to Chat Generative Pre-trained Transformer, and the answer given by the chatbot.

* Corresponding author. Tel: +41 79 556 82 05, Email: stephane.fournier@chuv.ch

© The Author(s) 2023. Published by Oxford University Press on behalf of the European Society of Cardiology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

The integration of scientific research, clinical studies, expertise, and patient values is vital for making informed medical decisions. ChatGPT's inability to provide medical information according to evidence-based medicine criteria, as stated by the authors, underscores the need for further development and evaluation.¹⁰ While large language model's limitations hinder its immediate adoption in evidence-based medicine, we believe that future advancements hold promise.

In conclusion, while we align with the concerns raised by the authors regarding ChatGPT's current limitations in evidence-based medicine and while we agree that caution, critical evaluation, and transparency are essential when considering the integration of AI systems like ChatGPT into clinical practice, we also believe that ChatGPT's performance is impressive especially given the fact this technology is extremely young. Accordingly, we remain optimistic about the future potential of AI in healthcare and believe that with continued research and development, specialized systems may eventually bridge the gaps and contribute to evidence-based medical decision-making.

Funding

None declared.

Conflict of interest: None declared.

Data availability

The data that support the findings of this study are available from the corresponding author upon request.

References

1. Haverkamp W, Cagnina A, Tennenbaum J, Strodthoff N. ChatGPT fails the test of evidence-based medicine? *Eur Heart J Digital Health* 2023.
2. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell* 2023; **6**:1169595.
3. Skalidis I, Cagnina A, Luangphiphat W, Mahendiran T, Muller O, Abbe E, et al. ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story? *Eur Heart J Digit Health* 2023;**4**:279–281.
4. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *medRxiv* 2023;**2**. 2022.12.19.22283643.
5. Salihi A, Gadiro M A, Skalidis I, Meier D, Auberson D, Thanou D, et al. Towards AI-assisted cardiology: a reflexion on the performance and limitations of using large language models in clinical decision making. *Eurointervention* 2023.
6. European Society of Cardiology. About the European Exam in Core Cardiology (EECC). [https://www.escardio.org/Education/Career-Development/European-Exam-in-Core-Cardiology-\(EECC\)](https://www.escardio.org/Education/Career-Development/European-Exam-in-Core-Cardiology-(EECC)) (10 July 2023).
7. Gordijn B, Have HT. ChatGPT: evolution or revolution? *Med Health Care Philos* 2023. Published online ahead of print 01/03/2023. PMID: 36656495.
8. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect Dis* 2023;**23**:405–406.
9. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet* 2017;**390**:415–423.
10. Itchhaporia D. Artificial intelligence in cardiology. *Trends Cardiovasc Med* 2022;**32**: 34–41.