ESC European Society of Cardiology

European Heart Journal - Digital Health (2023) **4**, 364–365
https://doi.org/10.1093/ehjdh/ztad042

LETTER TO THE EDITOR

Performance of artificial intelligence in answering cardiovascular textual questions

Ioannis Skalidis , Aurelien Cagnina, and Stephane Fournier *

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

Online publish-ahead-of-print 17 July 2023

Commentary article to: 'Does ChatGPT succeed in the European Exam in Core Cardiology?', by C. Plummer *et al.* https://doi.org/10.1093/ehjdh/ztad040.

We would like to thank Plummer *et al*. on our publication examining the performance of Chat Generative Pre-trained Transformer (GPT) in the European Exam in Core Cardiology (EECC).[1,2] Evaluating the limitations and potential biases in the study is essential.[3] In this response, we aim to assess the concerns raised by Plummer *et al*. while discussing ChatGPT's performance in the context of the EECC, a comprehensive European board examination in core cardiology.[4,5]

Plummer *et al.* rightly points out that the structure, format, and editorial process of the questions used in the study may not exactly mirror those of the EECC. While the original questions are confidential and without having access to them, the questions used were derived from the official mock exams from the European Society of Cardiology (same committee that finally makes the questions of the exam), which are designed to be similar in difficulty and complexity to the actual exam. It is important to acknowledge that while the questions may not be identical, the questions from the mock exam as well as from the traditional preparation material by Brauwnwald and StudyPRN are the closest in structure that the candidates have access to.[6,7]

We concur with the author's observation that ChatGPT's inability to handle image and video-based questions is a significant limitation. The EECC encompasses a wide range of medical imaging modalities questions, including coronary angiography, electrocardiograms, echocardiography, pacemaker tracings, etc. As ChatGPT's capability is currently limited to text-based inputs, it is unable to attempt approximately 30% of the questions. However, if we consider the theoretical possibility of ChatGPT selecting answers for these image or video-based questions at random, there is a 20% chance (given the format of five-option multiple choice questions) that it would choose the correct answer. This could potentially represent seven correct responses (20% of 36). For the text-based questions, the maximum achievable score would be 84 out of 120 (70%). Given ChatGPT's overall accuracy of 58.8% in these types of questions, it is projected that it would provide 49 correct responses. With the addition of the potential seven correct responses from the image or video-based questions, it could potentially deliver a total of 56 correct responses. This total represents an overall performance of approximately 47% which falls just short of the pass mark for the 2023 edition of the exam, which was set slightly lower than previous years, at 52%. Thus, while ChatGPT did not meet the threshold in this instance, it narrowly missed passing the EECC—an impressive feat considering the very young age of this tool.

We acknowledge the concern raised by Plummet *et al*. regarding the answer format employed in the study. However, it is important to clarify that the questions used in the study, derived from official mock exams, Braunwald, and StudyPRN databases, followed the same structure as the original exam. Thus, they were presented as multiple choice questions with a single correct answer among five possible options. The answer among the options by ChatGPT was compared to the correct answer defined by the respective exam preparation material, for the specific multiple choice question. In that sense, the format of the questions used was the same to the exam's questions.

In the broader perspective, it is crucial to note that the purpose of evaluating ChatGPT's performance in a setting like the EECC extends beyond merely determining whether the artificial intelligence (AI) can pass a multiple choice exam. Given the rigorous quality control and security measures in place, the use of such tools by students during examinations is highly unlikely and ethically impermissible. More importantly, we seek to understand the potential of ChatGPT and other large language models to assist medical professionals in their practice. The true value of these models lies in their potential ability to handle a vast amount of medical knowledge, process it rapidly, and provide useful responses to textual queries. While our study provides initial insights into the capabilities of ChatGPT within the confines of an exam-like structure, its real-world application in aiding clinicians by providing succinct information, summarizing complex medical literature, or supporting decision-making processes is the ultimate goal. However, it is essential to move forward with caution. While the prospect of AI in healthcare is exciting, we must remain vigilant about the challenges, ethical implications, and potential for misuse.[8]

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. Plummer C, Mathysen D, Lawson C. Does ChatGPT succeed in the EECC? *Eur Heart J Digit Health* 2023;**4**:279–281.

* 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

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. Plummer C, Bowater S, Hall J, Lawson C, Ooues G, Price S, *et al.* Behind the scenes of the European examination in general cardiology. *Heart* 2019;**105**: 889–890.
5. Plummer C, Fox K, Borg A, Sionis A, Burri H, Naka K. ESC Congress 2018. Take the Challenge! The European Exam on General Cardiology (EEGC). https://esc365.escardio.org/session/24165 (26 August 2018)
6. StudyPRN. Cardiology EECC (Free Trial). https://www.studyprn.com/. (25 June 2023).
7. Lilly LS. *Braunwald's heart disease review and assessment*. In: A Companion to Braunwald's Heart Disease. 12th ed. Amsterdam: Elsevier; 2022.
8. Salihu A, Gadiri 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.