ESC European Society of Cardiology

European Heart Journal - Digital Health (2023) **4**, 362–363
https://doi.org/10.1093/ehjdh/ztad040

LETTER TO THE EDITOR

Does ChatGPT succeed in the European Exam in Core Cardiology?

Chris Plummer [1,2,3,*], Danny Mathysen[4], and Clive Lawson[5]

[1]Department of Cardiology, Freeman Hospital, The Newcastle upon Tyne Hospitals NHS Foundation Trust, Freeman Road, High Heaton, Newcastle upon Tyne, NE7 7DN, UK; [2]European Society of Cardiology, The European Heart House, 2035 Route des Colles, Les Templiers - CS 80179 Biot, 06903 Sophia Antipolis, France; [3]The Union of European Medical Specialists Cardiology Section, 24 Rue de l'Industrie, 1040 Brussels, Belgium; [4]Faculty of Medicine and Health Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium; and [5]Department of Cardiology, Tunbridge Wells Hospital, Tonbridge Road, Pembury, Tunbridge Wells, Kent TN2 4QJ, UK

Online publish-ahead-of-print 16 July 2023

Commentary article to: 'Performance of artificial intelligence in answering cardiovascular textual questions', by I. Skalidis *et al*. https://doi.org/10.1093/ehjdh/ztad042.

We read with great interest the study by Skalidis *et al.*,[1] *ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story?*, which reports on the performance of Chat Generative Pre-trained Transformer (ChatGPT), an artificial intelligence (AI) chatbot developed by OpenAI, on textual questions in the style of the European Examination in Core Cardiology (EECC).[2] The ability of AI systems to answer structured medical questions is the subject of active research and is of considerable interest both in medical education and the delivery of clinical care, but it is important to acknowledge the limitations of current technology and of the ways in which its accuracy is tested.

The EECC comprises 120 'Best-of Five' multiple choice questions in English delivered online with remote proctoring over 3 hours, once each year. The questions are based on European Society of Cardiology (ESC) clinical practice guidelines[3] and the ESC CardioMed textbook.[4] They cover the broad range of the ESC core curriculum[5] at the level appropriate for independent practice. The EECC is used by 28 national cardiology societies and 2 international cardiology partners as a knowledge assessment for cardiology trainees. Each question is written by an experienced cardiologist, edited by a group of their peers in a standard format, reviewed by the question selection and standard setting groups, and undergoes final review by the chairs of the exam board and standard setting group before it can appear in an exam. All questions have short clinical stems, of which 30% include an image or video clip, a short question, and five possible answers from which candidates are required to choose the most appropriate. There are no 'negative questions' asking which option is incorrect or least likely, and there are no questions which are independent of the clinical scenario. The performance of each question in the EECC is reviewed by an independent psychometrician before it is used in the calculation of candidates' scores. The pass mark is determined using the Hofstee method[6] based on the performance of candidates in the delivered exam. In recent years, the EECC pass mark has varied between 65 (54%) and 70 (58%) correct responses out of 120 questions.

The questions used to test the performance of ChatGPT in this study are not included in the publication, so it is not possible to compare them to current EECC standards. The authors state that 'A total of 488 publicly-available single-answer MCQs were randomly obtained' from 'sample exam questions released since 2018 from the official ESC website, as well as the 2022 edition of StudyPRN and Braunwald's Heart Disease Review and Assessment (BHDRA)'.[1] Some of those available on the ESC website, for example those used in an ESC congress session in 2018,[7] were written before the EECC's realignment to the ESC core curriculum in 2020 and were not subject to EECC editorial review. None of the questions in the StudyPRN Cardiology EECC (Free Trial)[8] is written in EECC style, and Braunwald's Heart Disease Review and Assessment[9] contains questions written to 'enable fellows, residents, and practitioners to prepare for board exams in cardiovascular medicine' in the USA so are not primarily aligned with ESC clinical practice guidelines.[3]

This study raises some important questions, but we would like to draw attention to some limitations of its conclusion 'that ChatGPT succeeds in the EECC'.[1]

(1) The structure, format, editorial process, and guideline-directed answers of at least some of the questions used in this study are significantly different to questions used in the EECC. ChatGPT has not answered any delivered EECC questions.
(2) ChatGPT would be unable to attempt 36 (30%) of the 120 questions in the EECC which include an image or video clip, making its maximum possible mark 84 (70%). Its overall accuracy of 58.8% in text-based questions would therefore be expected to deliver a mark of 49 correct responses (41%) which would be below any previous pass mark. It was most successful for StudyPRN, where it indicated the correct answer in 63.8% of questions, but again, this would result in only 54 correct responses (45%).
(3) ChatGPT appears to have been asked for the answer to the question, rather than to select from the five potential answers, with the authors then determining whether this was correct. This is a different process to that required of EECC candidates and suggests that ChatGPT would not be able to answer EECC questions without human interpretation.

The evidence presented suggests that ChatGPT in its current form would not be able to pass the EECC. This study does, however, raise the possibility that it could significantly increase the likelihood of a candidate's success in this or similar exams. The EECC's remote proctoring

* Corresponding author. Email: Chris.Plummer@nhs.net

prevents candidates from collaborating, or consulting physical or online resources, so ChatGPT would not be able assist a candidate in the EECC. It does, however, raise serious questions about the use of open book exams in the era of highly advanced large language model AI.

Funding

No funding was used for the completion of this work.

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. 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.
2. European Society of Cardiology. European Exam in Core Cardiology (EECC). https://www.escardio.org/Education/Career-Development/European-Exam-in-Core-Cardiology-(EECC) (25 June 2023).
3. European Society of Cardiology. Clinical practice guidelines. https://www.escardio.org/Guidelines/Clinical-Practice-Guidelines (25 June 2023).
4. Camm AJ, Lüscher TF, Maurer G, Serruys PW, eds. *The ESC Textbook of Cardiovascular Medicine*, 3rd ed, The European Society of Cardiology Series, Oxford: ESC Publications; 2018. https://doi.org/10.1093/med/9780198784906.001.0001 (25 June 2023).
5. Tanner FC, Brooks N, Fox KF, Gonçalves L, Kearney P, Michalis L, *et al.* ESC core curriculum for the cardiologist. *Eur Heart J* 2020;**41**:3605–3692.
6. 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.
7. Take the Challenge! The European Exam on General Cardiology (EEGC) Chris Plummer, Kevin Fox, Alexander Borg, Alessandro Sionis, Haran Burri and Katerina Naka. ESC Congress 2018. 26 August 2018. https://esc365.escardio.org/session/24165 (25 June 2023).
8. StudyPRN Cardiology EECC (Free Trial). https://www.studyprn.com/ (25 June 2023).
9. Lilly LS. *Braunwald's Heart Disease Review and Assessment, 12th ed,* A Companion to Braunwald's Heart Disease, Oxford: Elsevier Limited; 2022, ISBN 9780323835138.