

GPT-4's Performance on the European Board of Interventional Radiology Sample Questions

Muhammed Said Beşler¹ 

Received: 6 March 2024 / Accepted: 8 March 2024 / Published online: 26 March 2024

© Springer Science+Business Media, LLC, part of Springer Nature and the Cardiovascular and Interventional Radiological Society of Europe (CIRSE) 2024

I read with great interest the article by Scheschenja et al., which investigates the feasibility of using generative pre-trained transformers (GPT) for patient education before interventional radiology procedures [1]. In this letter, I would like to offer a perspective on the capability of the multimodal large language model (LLM) to assess clinical case scenarios.

ChatGPT language models hold significant potential in interventional radiology, from patient education to serving as a medical decision support tool [2]. Particularly, the latest version, GPT-4, has shown high performance in various radiology board examinations [3, 4]. GPT-4, as a multimodal LLM, can provide text-based answers to image-based inputs. I examined 31 sample clinical case scenario questions from the European Board of Interventional Radiology (EBIR) examination made available by the Cardiovascular and Interventional Radiology Society of Europe, consisting of 7 cases (https://www.cirse.org/wp-content/uploads/2023/07/EBIR-Sample-Questions-clinical-case-scenario_-English.pdf). Each case comprised a series of related multiple-choice questions, with either 4 or 5 options. All responses were obtained using GPT-4 in March 2024. In questions with multiple correct answers, an accuracy of 88.9% (8 of 9) was achieved, and for those with a single correct answer, an accuracy of 72.7% (16 of 22) was observed. For questions containing only text, the accuracy was 90.9% (10 of 11), and for those with medical images, it was 60% (12 of 20). Overall, a correct answer

performance of 71% (22 of 31) was demonstrated. The potential for multimodal LLMs to pass a full EBIR examination is a subject worth exploring. Further studies are needed to demonstrate their potential to assist in clinical cases in interventional radiology.

Funding This study was not supported by any funding.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent For this type of study informed consent is not required.

Consent for Publication For this type of study consent for publication is not required.

References

1. Scheschenja M, Viniol S, Bastian MB, Wessendorf J, König AM, Mahnken AH. Feasibility of GPT-3 and GPT-4 for in-depth patient education prior to interventional radiological procedures: a comparative analysis. *Cardiovasc Intervent Radiol*. 2024;47(2):245–50. <https://doi.org/10.1007/s00270-023-03563-2>.
2. Campbell WA 4th, Chick JFB, Shin D, Makary MS. Understanding ChatGPT for evidence-based utilization in interventional radiology. *Clin Imaging*. 2024. <https://doi.org/10.1016/j.clinimag.2024.110098>.
3. Sood A, Mansoor N, Memmi C, Lynch M, Lynch J. Generative pretrained transformer-4, an artificial intelligence text predictive model, has a high capability for passing novel written radiology exam questions. *Int J Comput Assist Radiol Surg*. 2024. <https://doi.org/10.1007/s11548-024-03071-9>.

✉ Muhammed Said Beşler
msbesler@gmail.com

¹ Department of Radiology, Kahramanmaraş Necip Fazıl City Hospital, 46050 Kahramanmaraş, Turkey

4. Ariyaratne S, Jenko N, Mark Davies A, Iyengar KP, Botchu R. Could ChatGPT pass the UK radiology fellowship examinations? *Acad Radiol.* 2023. <https://doi.org/10.1016/j.acra.2023.11.026>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.