ARCHIVES OF THE TURKISH SOCIETY OF CARDIOLOGY

Official journal of the TURKISH SOCIETY OF CARDIOLOGY

Comparative Evaluation of Chatbot Responses on Coronary Artery Disease

Koroner Arter Hastalığına İlişkin Chatbot Yanıtlarının Karşılaştırmalı Değerlendirilmesi

LETTER TO THE EDITOR
EDİTÖRE MEKTUP

To the Editor,

We would like to comment on the article titled "Comparative Evaluation of Chatbot Responses on Coronary Artery Disease".[1] Pay et al.[1] assessed the accuracy and reproducibility of responses from several models, including ChatGPT and Bing, to frequently asked questions about coronary artery disease. They found that ChatGPT demonstrated the highest accuracy and reproducibility.[1] This study provides valuable insights into the performance of various natural language processing (NLP) chatbots in addressing common inquiries related to coronary artery disease (CAD). The involvement of cardiologists in independently rating the responses added an important layer of clinical expertise, ensuring that assessments of accuracy and completeness were medically relevant. The findings indicate that ChatGPT outperforms Gemini and Bing in terms of providing complete and accurate responses, with notable differences in memory capabilities. However, there are still areas that require further exploration and improvement.

Firstly, although ChatGPT exhibited the highest accuracy, the study did not explore the specific reasons behind the performance differences among the NLP models. Are these variations in response quality due to differences in underlying algorithms, training data, or the way each chatbot interacts with the questionnaire? Investigating the influence of training data diversity, particularly the range of medical resources used, could shed light on chatbot performance.[2] Additionally, understanding the specific algorithms employed by each model might help clarify the performance disparities observed. Further research into how these models interpret medical terminology and clinical context could contribute to their future refinement. Additionally, the study could benefit from exploring the potential influence of external factors, such as the model's ability to update their knowledge bases in response to new medical data or clinical guidelines. Furthermore, the study did not address the importance of interpretability in clinical practice. Even if a chatbot provides accurate responses, healthcare professionals may struggle to trust or apply the information if they do not understand the reasoning behind the answers.

Reproducibility concerns are equally significant in the context of medical decision-making. Although the study shows that ChatGPT achieved the highest reproducibility scores, the variations across platforms raise concerns about the consistency of chatbot responses. Can clinicians rely on chatbots to provide consistent answers? Future research should consider these challenges, especially in dynamic clinical environments where patient data may change frequently, and in domains such as CAD, where accuracy is critical. Strategies to enhance response consistency across platforms should also be explored.

In terms of innovation and future directions, one promising area of exploration is the integration of chatbots with real-time patient data to deliver more personalized and actionable insights. The research could also be extended by evaluating chatbot performance across a broader range of medical specialties, as well as assessing their effectiveness in various languages and cultural contexts. It may also be valuable to explore the potential of chatbots to integrate with other healthcare tools, such as electronic health records, to support clinical decision-making. Additionally, NLP

Hinpetch Daungsupawong[1]
Viroj Wiwanitkit[2]

[1]*Private Academic Consultant, Phonhong, Lao People's Democratic Republic*
[2]*Department of Community Medicine, D Y Patil Medical College, Hospital and Research Centre, Dr. DY Patil Vidyapeeth University, Pune, India*

Corresponding author:
Hinpetch Daungsupawong
✉ hinpetchdaung@gmail.com

Received: January 14, 2025
Accepted: February 08, 2025

Cite this article as: Daungsupawong H, Wiwanitkit V. Comparative Evaluation of Chatbot Responses on Coronary Artery Disease. *Turk Kardiyol Dern Ars.* 2025;53(5):370-371.

DOI: 10.5543/tkda.2025.66811

chatbots could be improved by incorporating feedback loops with healthcare professionals, enabling continuous refinement based on real-world use and clinical feedback.

Conflict of Interest: The authors have no conflicts of interest to declare.

Funding: The authors declared that this study received no financial support.

References

1. Pay L, Yumurtaş AÇ, Çetin T, Çınar T, Hayıroğlu Mİ. Comparative Evaluation of Chatbot Responses on Coronary Artery Disease. *Turk Kardiyol Dern Ars*. 2025;53(1):35–43. [CrossRef]
2. Berger L, Haberbusch M, Moscato F. Generative adversarial networks in electrocardiogram synthesis: Recent developments and challenges. *Artif Intell Med*. 2023;143:102632. [CrossRef]