ARCHIVES OF THE
TURKISH SOCIETY
OF CARDIOLOGY

Official journal of the
TURKISH
SOCIETY OF
CARDIOLOGY

Reply to the Letter to the Editor: "Comparative Evaluation of Chatbot Responses on Coronary Artery Disease"

Editöre Mektuba Yanıt: "Koroner Arter Hastalığında Sohbet Botu Yanıtlarının Karşılaştırmalı Değerlendirmesi"

LETTER TO THE EDITOR REPLY
EDİTÖRE MEKTUP YANITI

To the Editor,

We would like to express our sincere gratitude to the authors[1] for their careful review and insightful comments regarding our work. We highly appreciate their constructive criticism, which provides a valuable opportunity to further enhance our study.

Firstly, our study identified significant performance differences among chatbots, with ChatGPT demonstrating the highest accuracy. However, we did not conduct a detailed technical analysis of underlying factors such as model architecture, training data, or interaction strategies. We acknowledge that the diversity and quality of training data, particularly the inclusion of medical sources, likely impact response quality. Due to the proprietary nature of these models, detailed information on their internal mechanisms is not publicly accessible. A recent study has shown that differences in chatbot responses may arise from variations in communication and language styles, which significantly influence users' perceptions of effectiveness, trustworthiness, and usability.[2] This represents an important area for future research to better understand and improve artificial intelligence (AI) tools in clinical practice.

Secondly, we fully acknowledge that even when a chatbot provides accurate information, healthcare professionals may find it difficult to rely on or apply such information if the underlying rationale is not transparent or well explained. Indeed, systematic reviews highlight that explainable AI and transparency are critical to fostering trust in healthcare AI systems.[3] However, the primary aim of our study was to evaluate and compare the factual accuracy and guideline concordance of chatbot responses to frequently asked questions related to coronary artery disease. Our focus was therefore on assessing informational correctness rather than explanatory depth or clinical interpretability. In addition, we agree that explainability, source referencing, and clinical reasoning are essential components of any AI tool intended for healthcare use, as they are key to building trust and supporting clinical decision-making. While our study demonstrates the potential of large language models to provide medically accurate responses, it also highlights the need for further research on their transparency and usability in clinical practice.

Thirdly, we agree that reproducibility and response consistency are essential for clinical trust, particularly in dynamic settings like coronary artery disease management, where patient data frequently change and decisions rely on stable information. Our study found that ChatGPT achieved the highest reproducibility scores. However, variability across different chatbot platforms poses a potential risk when relying on these tools for consistent clinical guidance. Consistent with our findings, Shiferaw et al.[4] demonstrated that a lack of response consistency in repeated ChatGPT queries poses a significant challenge to its reliability in healthcare settings, underscoring the importance of reproducibility for clinical trust. This highlights an important area for further investigation. Future research should focus on how factors such as model prompts, session resets, or minor rephrasings influence response consistency, as well as the impact of platform-specific design choices on reproducibility. Developing strategies to improve and standardize reliability across chatbot platforms will be critical for their successful integration into clinical workflows.

Levent Pay[1]
Ahmet Çağdaş Yumurtaş[2]
Tuğba Çetin[3]
Tufan Çınar[4]
Mert İlker Hayıroğlu[3]

[1]*Department of Cardiology, İstanbul Haseki Training and Research Hospital, İstanbul, Türkiye*
[2]*Department of Cardiology, Kars Harakani State Hospital, Kars, Türkiye*
[3]*Department of Cardiology, Dr. Siyami Ersek Thoracic and Cardiovascular Surgery Training Hospital, İstanbul, Türkiye*
[4]*Department of Medicine, University of Maryland Medical Center Midtown Campus, Maryland, USA*

Corresponding author:
Levent Pay
✉ leventpay@hotmail.com

Cite this article as: Pay L, Yumurtaş AÇ, Çetin T, Çınar T, Hayıroğlu Mİ. Reply to the Letter to the Editor: "Comparative Evaluation of Chatbot Responses on Coronary Artery Disease". *Turk Kardiyol Dern Ars.* 2025;53(5):372-373.

DOI: 10.5543/tkda.2025.05691

Lastly, we fully agree that integrating chatbots with real-time patient data and electronic health records, alongside expanding evaluations across languages, specialties, and cultural contexts, is a crucial path forward. Establishing feedback loops with healthcare professionals will also be essential to align these tools with real-world clinical needs. We consider these steps important to making AI-assisted tools more actionable, personalized, and clinically relevant.

References

1. Daungsupawong H, Wiwanitkit V. Comparative Evaluation of Chatbot Responses on Coronary Artery Disease. *Turk Kardiyol Dern Ars*. 2025;53(5):370-371.
2. Koscelny SN, Sadralashrafi S, Neyens DM. Generative AI responses are a dime a dozen; Making them count is the challenge - Evaluating information presentation styles in healthcare chatbots using hierarchical Bayesian regression models. *Appl Ergon*. 2025;128:104515. [CrossRef]
3. Eke CI, Shuib L. The role of explainability and transparency in fostering trust in AI healthcare systems: a systematic literature review, open issues and potential solutions. *Neural Comput & Applic*. 2025;37(4):1999-2034. [CrossRef]
4. Shiferaw MW, Zheng T, Winter A, Mike LA, Lingtak Chan LN. Assessing the accuracy and quality of artificial intelligence (AI) chatbot-generated responses in making patient-specific drug-therapy and healthcare-related decisions. *BMC Med Inform Decis Mak*. 2024;24(1):404. [CrossRef]