ARCHIVES OF THE
TURKISH SOCIETY
OF CARDIOLOGY

Official journal of the TURKISH SOCIETY OF CARDIOLOGY

Can Large Language Models Guide Aortic Stenosis Management? A Comparative Analysis of ChatGPT and Gemini AI

Büyük Dil Modelleri Aort Darlığı Yönetimine Rehberlik Edebilir mi? ChatGPT ve Gemini Yapay Zekanın Karşılaştırmalı Analizi

LETTER TO THE EDITOR
EDİTÖRE MEKTUP

To the Editor,

I read Sezgin et al.'s article titled "Can Large Language Models Guide Aortic Stenosis Management? A Comparative Analysis of ChatGPT and Gemini AI," published in Archives of the Turkish Society of Cardiology, with great interest.[1] The study provides valuable insight into the use of widely accessible large language models (LLMs) in the management of severe aortic stenosis. I would like to share a few constructive observations that may be useful for future research or manuscript refinement.

Large language models, widely used across medical and non-medical domains, are becoming increasingly integrated into everyday life.[2] These models have also become prominent in contemporary medical research due to their potential to support both medical education and clinical decision-making processes. However, when ethical considerations and professional evaluation standards are taken into account, a controlled and supervised use of LLMs appears to be the most appropriate approach.[3] In the context of LLM use, question phrasing is crucial; even slight variations in wording can result in significantly different answers. As prompts become more specific, detailed, and clearly formulated, the responses from LLMs tend to become more accurate, consistent, and aligned with the intended objective.[4]

In this study,[1] each question was asked independently in a separate session, preventing the model from retaining memory of previous interactions. This approach eliminated the risk of biased responses that could arise from content being carried over from earlier questions. We also examined the questions that contributed to this difference. We believe that some of the information provided may be insufficient for both guideline-based evaluation and real-world clinical judgment, particularly in the clinical scenario questions. Question 33 – "Is surgery indicated in a 59-year-old asymptomatic patient with left ventricular (LV) hypertrophy and ejection fraction (EF) 52%?" – does not explicitly refer to aortic stenosis, potentially reducing the question's alignment with the intended evaluation scope. When we tested ChatGPT 5.1 (the current version) by presenting the same question in two different formats, the model generated two distinct answers (Figure 1 and 2). Question 35 – "How should combined treatment be planned for a 64-year-old patient with 90% left anterior descending (LAD) artery stenosis and severe aortic stenosis (AS)?" – appears to be phrased too broadly. In our opinion, the optimal management strategy would depend on additional information such as the anatomical location of the LAD lesion, further echocardiographic parameters, surgical risk stratification, and relevant comorbidities, none of which are specified in the scenario. Although the study was conducted using a general evaluation framework, it is essential to further refine and expand the clinical scenario questions to ensure better alignment with real-world clinical decision-making and patient variability. Although the clinical scenarios lack detailed contextual information, the study still provides valuable insight, particularly in demonstrating Gemini's near-perfect performance.

Mehmet Uğur Çalışkan[1]

Nezaket Merve Yaman Akkuş[2]

[1]Department of Cardiology, Amasya Training and Research Hospital, Amasya, Türkiye
[2]Department of Cardiology, Sincan Training and Research Hospital, Ankara, Türkiye

Corresponding author:
Mehmet Uğur Çalışkan
✉ ugurkobian@gmail.com

Received: November 25, 2025
Accepted: December 03, 2025

Cite this article as: Çalışkan MU, Yaman Akkuş NM. Letter to the Editor: Can Large Language Models Guide Aortic Stenosis Management? A Comparative Analysis of ChatGPT and Gemini AI. *Turk Kardiyol Dern Ars.* 2026;54(1):83–84.

DOI: 10.5543/tkda.2025.03388

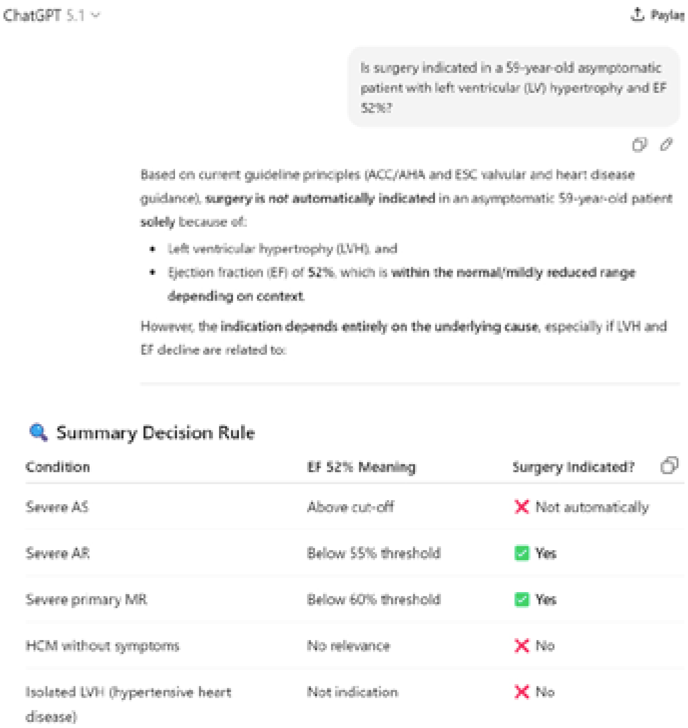

Figure 1. ChatGPT 5.1 response to question 33.

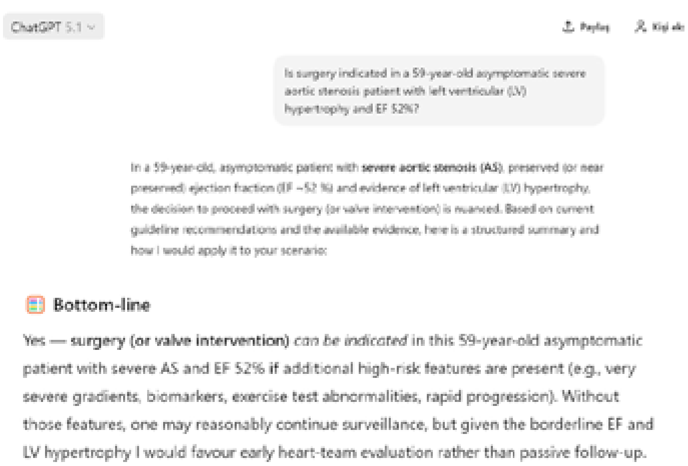

Figure 2. ChatGPT response to question 33 when aortic stenosis is specified.

In conclusion, despite its limitations, the work conducted by Sezgin et al. provides an important foundation for future research exploring the role of large language models in aortic stenosis. Future studies are needed to explore the decision-making mechanisms of LLMs in patients with aortic stenosis, particularly using real-world clinical scenarios.

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. Sezgin A, Tanık VO, Akdoğan M, et al. Can Large Language Models Guide Aortic Stenosis Management? A Comparative Analysis of ChatGPT and Gemini AI. *Turk Kardiyol Dern Ars*. 2025;53(8):599-606. [CrossRef]
2. Yu E, Chu X, Zhang W, et al. Large Language Models in Medicine: Applications, Challenges, and Future Directions. *Int J Med Sci*. 2025;22(11):2792-2801. [CrossRef]
3. Shool S, Adimi S, Saboori Amleshi R, Bitaraf E, Golpira R, Tara M. A systematic review of large language model (LLM) evaluations in clinical medicine. *BMC Med Inform Decis Mak*. 2025;25(1):117. [CrossRef]
4. Meskó B. Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *J Med Internet Res*. 2023;25:e50638. [CrossRef]