



# Reply: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images

Alperen Elek<sup>1</sup>  
 Duygu Doğa Ekizalioğlu<sup>2</sup>  
 Ezgi Güler<sup>2</sup>

<sup>1</sup>Ege University Faculty of Medicine, İzmir, Türkiye

<sup>2</sup>Ege University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

## KEYWORDS

Abdomen, artificial intelligence, diagnostic imaging, large language models, magnetic resonance imaging, multidetector computed tomography

## Dear Editor,

We would like to express our gratitude to Daungsupawong and Wiwanitkit<sup>1</sup> for their insightful comments on our manuscript, "Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images"<sup>2</sup>. Their feedback is invaluable, and we appreciate the opportunity to engage in this discussion.

Our original study<sup>2</sup> demonstrated that as tasks become more complex, the model's performance significantly declines, particularly in predicting pathology. Although we acknowledge the current limitations of Bing in distinguishing imaging modalities and detecting contrast agents, we believe these challenges are inherent in the early stages of integrating artificial intelligence (AI) into medical imaging. Expecting AI technologies such as Bing with ChatGPT-4 to match the accuracy of radiologists from the outset may be overly ambitious. Nevertheless, it is worth considering that the fine-tuning or development of models specifically designed for diagnostic purposes might lead to improved results. Furthermore, enabling the upload of DICOM-format files, rather than single slices, could enhance performance.

Regarding the study's sample size, although we agree that a larger dataset would provide more robust insights, it is also important to recognize the value of small-scale, targeted studies in shaping future research directions. Preliminary evaluations can identify key areas for improvement and refine methodologies for larger trials, potentially leading to more focused and impactful studies.

The noted differences between Bing's integrated ChatGPT-4 and the standalone version further highlight the complexity of deploying AI models across various platforms. This variation reflects not only differences in implementation but also the broader challenge of ensuring consistent performance in diverse clinical environments. Addressing these nuances is critical to the safe and effective integration of AI into medical practice.

In conclusion, although our study explores the potential applications of Bing with ChatGPT-4, we stress the importance of using such tools as supplementary aids under expert supervision. They are far from replacing radiologists or being independently reliable for complex medical tasks. We hope this reply contributes to the ongoing discourse on the role of AI in diagnostic radiology and encourages the development of more robust, domain-specific solutions.

## Conflict of interest disclosure

The authors declared no conflicts of interest.

Corresponding author: Ezgi Güler

E-mail: gulerezgi@yahoo.com

Received 02 November 2024; revision requested 10 November 2024; accepted 16 December 2024.



Epub: 13.01.2025

Publication date:

DOI: 10.4274/dir.2024.243123

## References

1. Daungsupawong H, Wiwanitkit V. Letter to the editor re: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)
2. Elek A, Ekizaliođlu DD, Güler E. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)