



Letter to the editor re: evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging

Hinpetch Daungsupawong¹

Viroj Wiwanitkit²

¹Private Academic Consultant, Phonhong, Lao
People's Democratic Republic

²Saveetha Medical College, Saveetha Institute of
Medical and Technical Sciences, Chennai, India

Dear Editor,

The study titled "Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance imaging" presents a novel approach to medical image analysis.¹ This research aims to evaluate the effectiveness of Microsoft Bing, enhanced with ChatGPT-4 technology, in interpreting abdominal computed tomography (CT) and magnetic resonance imaging (MRI) data. Eighty abdominal images, including 44 CT and 36 MRI scans, were examined, and Bing's assessment was compared with that of a professional radiologist. The results showed that Bing could correctly identify CT scans with 95.4% accuracy and MRIs with 86.1% accuracy. However, Bing experienced some problems: wrongly identifying some images and poorly detecting anatomical regions, imaging planes, MRI sequences, and contrast agents. Bing discovered anomalies in only 35% of the images, with a 10.7% accuracy rate.

Bing's analysis suffers from inaccuracies in detecting imaging types, as evidenced by wrongly labeled CT and MRIs. The identification of MRI sequences and contrast agents was also poor, with success rates of 68.75% and 64.2%, respectively. Furthermore, Bing's low correct interpretation rates for anomalies underscore the difficulties of obtaining therapeutically useful information. Such limitations highlight its reliance on massive datasets and complex algorithms, which may not detect the tiny diagnostic signals found in medical imaging.

The study's comparative and descriptive design may limit its ability to address modest changes in image context or patient pathology. The sample size, although large, may be insufficient to draw broad conclusions. Bing's performance is context-dependent, and using only 80 photos may limit insights into its suitability for a wide range of clinical circumstances. Furthermore, the absence of real-time adaptive learning from feedback may impede the tool's progress, reducing its long-term relevance in radiology.

While Microsoft Bing incorporates ChatGPT-4 technology, there is evidence to suggest that its performance may not be as accurate or contextually aware as the standalone ChatGPT platform. This variation could be due to variances in how each system is taught and optimized for specific tasks. The standalone ChatGPT platform benefits from tailored training on various datasets, which improves performance in delivering nuanced and contextually relevant responses. OpenAI recently added memory features to its ChatGPT platform, allowing it to remember information between sessions for specific users.² As a result, when examining each system's usefulness in medical image analysis and other complicated domains, it is critical to consider its distinct strengths and limits.

To enhance Bing's diagnostic capabilities, future initiatives should focus on integrating more comprehensive datasets, encompassing a wider array of diseases, imaging modalities, and patient demographics. Language disparities in patient demographics dependent on the study location may have a major impact on the interpretation of the results.³ Continuous training with advanced deep learning techniques could further improve its ability to distinguish between various types of images and detect subtle anomalies. Investing in a

KEYWORDS

AI, audit, Bing, evaluation, radiology

Corresponding author: Hinpetch Daungsupawong

E-mail: hinpetchdaung@gmail.com

Received 20 August 2024; revision requested 09
September 2024; accepted 25 September 2024.



Epub: 21.10.2024

Publication date: 08.07.2025

DOI: 10.4274/dir.2024.242994

You may cite this article as: 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.* 2025;31(4):329-330.

real-time feedback loop in which Bing learns from radiologists' accurate diagnoses can help improve diagnostic accuracy. As Elek⁴ points out, the way a question is phrased to models like ChatGPT is critical to improving answer accuracy. Enabling web access in ChatGPT or seeking references from the PubMed database after asking queries may improve the model's accuracy.³ Finally, collaboration with medical practitioners might lead to improvements that address specific clinical needs. This will eventually make artificial intelligence systems like Bing more

reliable as a supplement to medical image analysis.

Footnotes

Conflict of interest disclosure

The authors declared no conflicts of interest.

References

1. Elek A, Ekizalioğlu DD, Güler E. Evaluating Microsoft Bing with ChatGPT-4 for the assessment of abdominal computed

tomography and magnetic resonance images. *Diagn Interv Radiol*. 2024. [\[CrossRef\]](#)

2. Zhong W, Guo L, Gao Q, Ye H, Wang Y. Memorybank: enhancing large language models with long-term memory. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2024;38(17):19724-19731. [\[CrossRef\]](#)
3. Elek A. The Role of large language models in radiology reporting. *AJR Am J Roentgenol*. 2023;221(5):707. [\[CrossRef\]](#)
4. Elek A. Improving accuracy in ChatGPT. *AJR Am J Roentgenol*. 2023;221(5):705. [\[CrossRef\]](#)