

# Artificial Intelligence in Diagnostic Radiology: Evaluating the Performance of ChatGPT on Radiology Core Examination



Bunnarin Theng<sup>1\*</sup>, BS; Joy Li<sup>1</sup>, BS; Roland Yu<sup>1</sup>, BS; Glenn Garcia<sup>1</sup>, MD

<sup>1</sup>University of Texas Medical Branch at Galveston, TX

\*butheng@utmb.edu

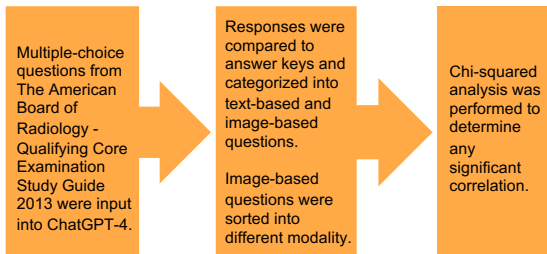
## Introduction

- Artificial Intelligence (AI) has shown promising advancements in recent years.
- ChatGPT is a powerful AI language model with great potential in medical practice and education, but its performance in radiology remains unclear<sup>1</sup>.
- ChatGPT currently has around 180.5 million users and the website generated 1.7 billion visits in October 2023<sup>2</sup>.

## Research Goal

The goal of this study is to assess the performance of ChatGPT's newest and most advanced version, GPT-4, on diagnostic radiology core examination questions with and without images and to explore its strengths and limitations.

## Methods



## Results

- ChatGPT-4 answered 56.8% of questions correctly (42 out of 74) and did not specifically answer 27.0% of the questions (20 out of 74).
- The model performed significantly better on text-based questions (76.9%, 20 out of 26) compared to image-based questions (45.8%, 22 out of 48) ( $P = 0.0197$ ).

**Table 1: Performance Summary of ChatGPT-4 on Radiology Core Examination Question**

	Correct	Incorrect	No Answer	Total	P-Value
<b>Text-Based Questions</b>	20 (76.9%)	5 (19.2%)	1 (3.8%)	26	0.02
<b>Image-Based Questions</b>	22 (45.8%)	15 (31.3%)	11 (22.9%)	48	
CT	6 (40.0%)	5 (33.3%)	4 (26.7%)	15	0.75
Radiograph	4 (36.4%)	4 (36.4%)	3 (27.3%)	11	
MRI/MRA	6 (60.0%)	4 (40.0%)	0 (0.0%)	10	
Ultrasound	4 (44.4%)	2 (22.2%)	3 (33.3%)	9	
Others (PET, V/Q, Thyroid)	2 (66.7%)	0 (0.0%)	1 (33.3%)	3	
<b>Total Questions</b>	42 (56.8%)	20 (27.0%)	12 (16.2%)	74	

## Conclusion

- ChatGPT-4 demonstrated notable performance in the radiology core examination, particularly excelling in text-based questions.
- However, its ability to process and respond to radiology image questions remains limited.
- The ever-changing field of artificial intelligence holds the potential to transform diagnostic radiology, providing invaluable support to radiologists in image interpretation.
- While ChatGPT-4 showcases the ability to recognize diverse image types, its accuracy is suboptimal and there is still room for more development and advancement in the future.

## Limitations and Future Direction

- The study utilized 74 practice questions from 2013 which may present a small and outdated sample size, potentially limiting the generalizability of its findings.
- Future studies can explore the potential role of ChatGPT in optimizing practicing radiologist workflow and the education of radiology trainees.

## References

- Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a Radiology Board-style Examination: Insights into Current Strengths and Limitations. *Radiology*. 2023;307(5):e230582. doi:10.1148/radiol.230582
- Exploding Topics. Number of ChatGPT Users. Available from: <https://explodingtopics.com/blog/chatgpt-users>