

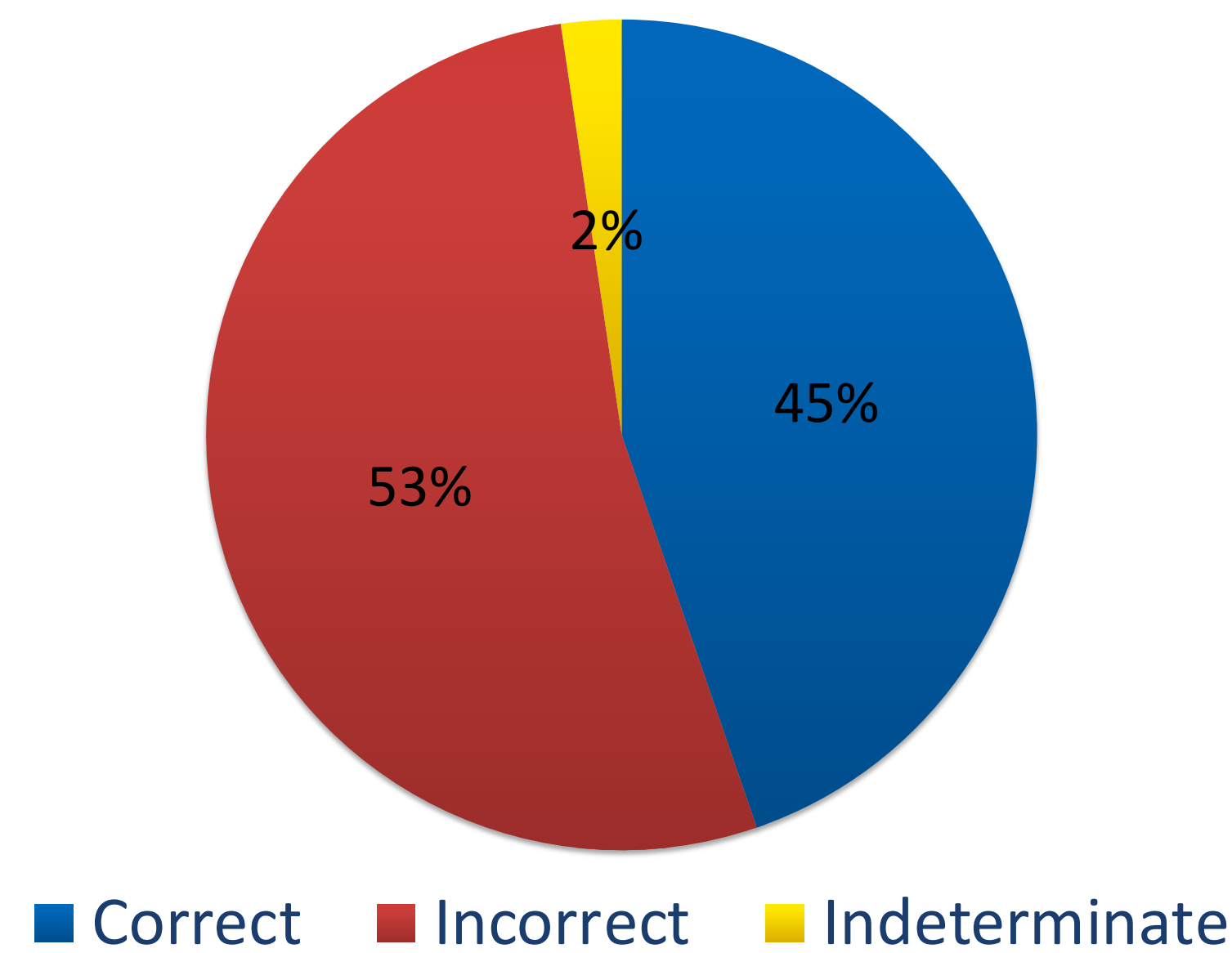
## Introduction

- Neural networks, deep learning, and artificial intelligence (AI) are quickly advancing technologies that are increasingly integrated into a spectrum of industries, including the field of medicine.
- Machine learning algorithms, neural networks, deep learning models, and natural language processing are all used to assist in creating treatment protocols, analyzing images, and improve documentation.<sup>1, 2</sup>
- ChatGPT (OpenAI, San Francisco, CA) is a new large language model launched by OpenAI in November 2022 programmed with reinforcement learning techniques.
- Notably, ChatGPT was able to pass the Step 1, Step 2, and Step 3 exams from the United States Medical Licensing Examination (USMLE).<sup>3</sup>
- This study aims to evaluate the AI technology's ability to answer otolaryngology board-style questions to assess its potential uses in medical education.

## Methods

- The version of ChatGPT used in this study was from the January 30th, 2023 release note.
- No specific priming or training was provided prior to the study.
- 500 otolaryngology board-style questions were submitted to ChatGPT from the Board Vitals (New York, NY) question bank.
- ChatGPT's performance was compared between the otolaryngology board-style questions and the USMLE questions.
- Additionally, the acceptability of its explanations were evaluated.
- The results were compared to that of the USMLE exams performed in a previous study using an unpaired T-test.<sup>3</sup>

## Results



**Chart 1: ChatGPT Raw Performance**

ChatGPT's performance on 500 ENT boards style questions. Indeterminant results were those where ChatGPT would make mistakes such as choosing multiple options.

### Ex. User Prompt

A 4 year old presents with a lateral neck mass. Computed tomography is performed showing a cystic tract anterior to the sternocleidomastoid muscle, passing between the internal and external branches of the carotid. If you are concerned about a fistula connected to the aerodigestive tract, where should you look for this tract's opening?

- A. Tonsillar bed
- B. Thyrohyoid membrane
- C. Pyriform sinus
- D. External auditory canal
- E. Nasopharynx

### ChatGPT Response

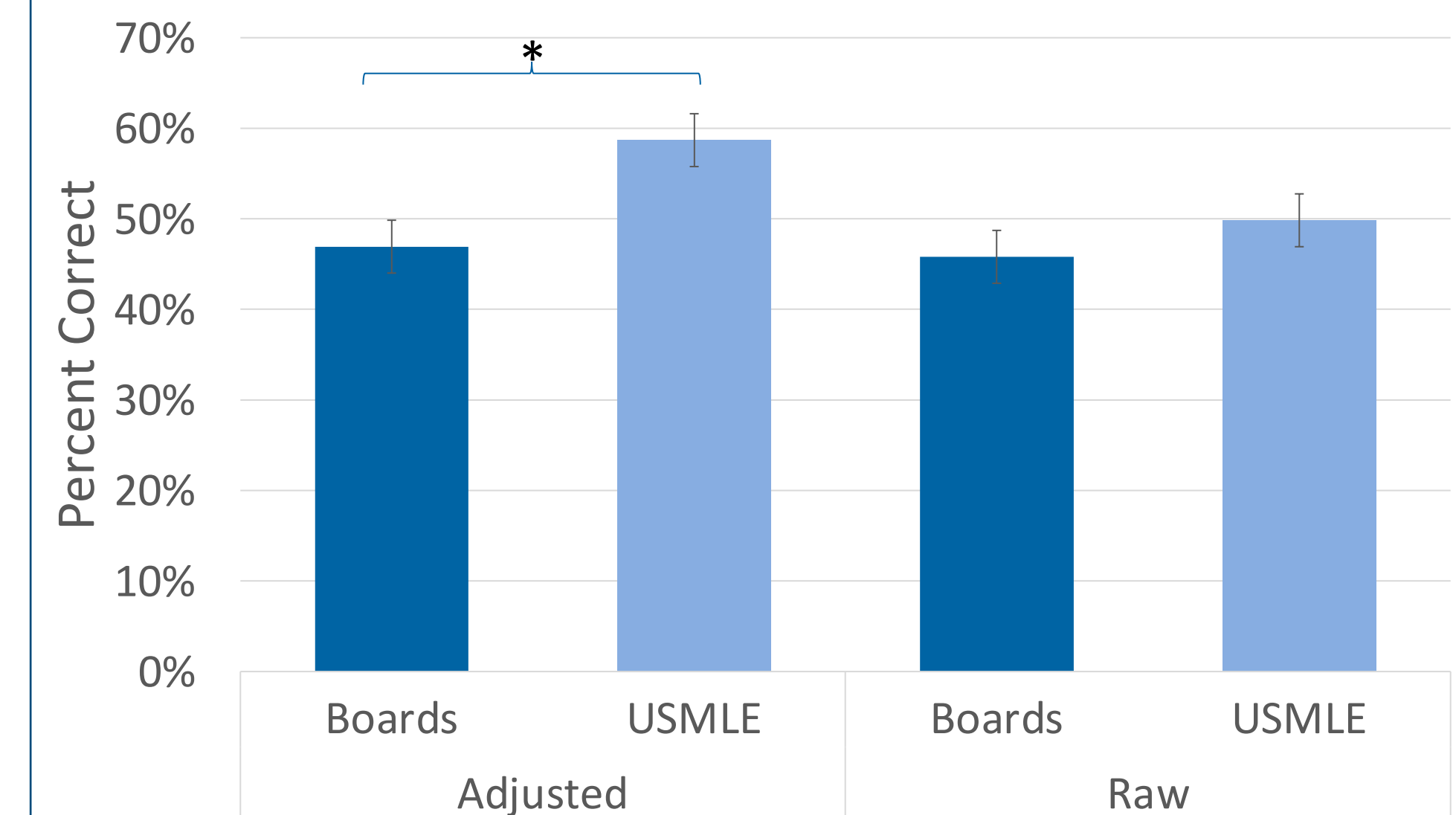
If you are concerned about a fistula connected to the aerodigestive tract in a 4-year-old with a lateral neck mass, you should look for the tract's opening in the option C: Pyriform sinus.

The pyriform sinus is a part of the hypopharynx located on each side of the larynx, between the thyroid cartilage and the cricoid cartilage. Fistulas between the aerodigestive tract and the lateral neck can occasionally occur, and the most common location for these fistulas is the pyriform sinus. Therefore, when investigating a lateral neck mass in this clinical scenario, the opening of the tract should be looked for in the pyriform sinus.

**Figure 1: Example Prompt and Response**

A response generated by ChatGPT to an example ENT boards style question.

## Results Continued



**Chart 2: ChatGPT ENT Boards vs USMLE**

ChatGPT's performance on the ENT boards style questions vs USMLE practice questions. The USMLE practice questions are a combination of the USMLE Step 1, 2, and 3. P value of the Adjusted data (excluding indeterminate results) was 0.001 while that of the Raw data was 0.267. \* denotes statistical significance.

## Conclusion

- ChatGPT performed at <50% accuracy on the otolaryngology boards-style question.
- ChatGPT performed significantly worse in the specialty-focused exam as opposed to the standardized exam administered to all physicians.
- The provided justification for its answer choices was also inaccurate.
- Further studies are needed to determine ChatGPT's role in the medical education system.

## References

1. Puschunder JM. The Potential for Artificial Intelligence in Healthcare. SSRN Electron J. 2020;6(2):94-98. doi:10.2139/ssrn.3525037
2. Sordo M. Introduction to neural networks in healthcare. Open Clin Doc. 2002.
3. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models. 2022;3786:1-25.