

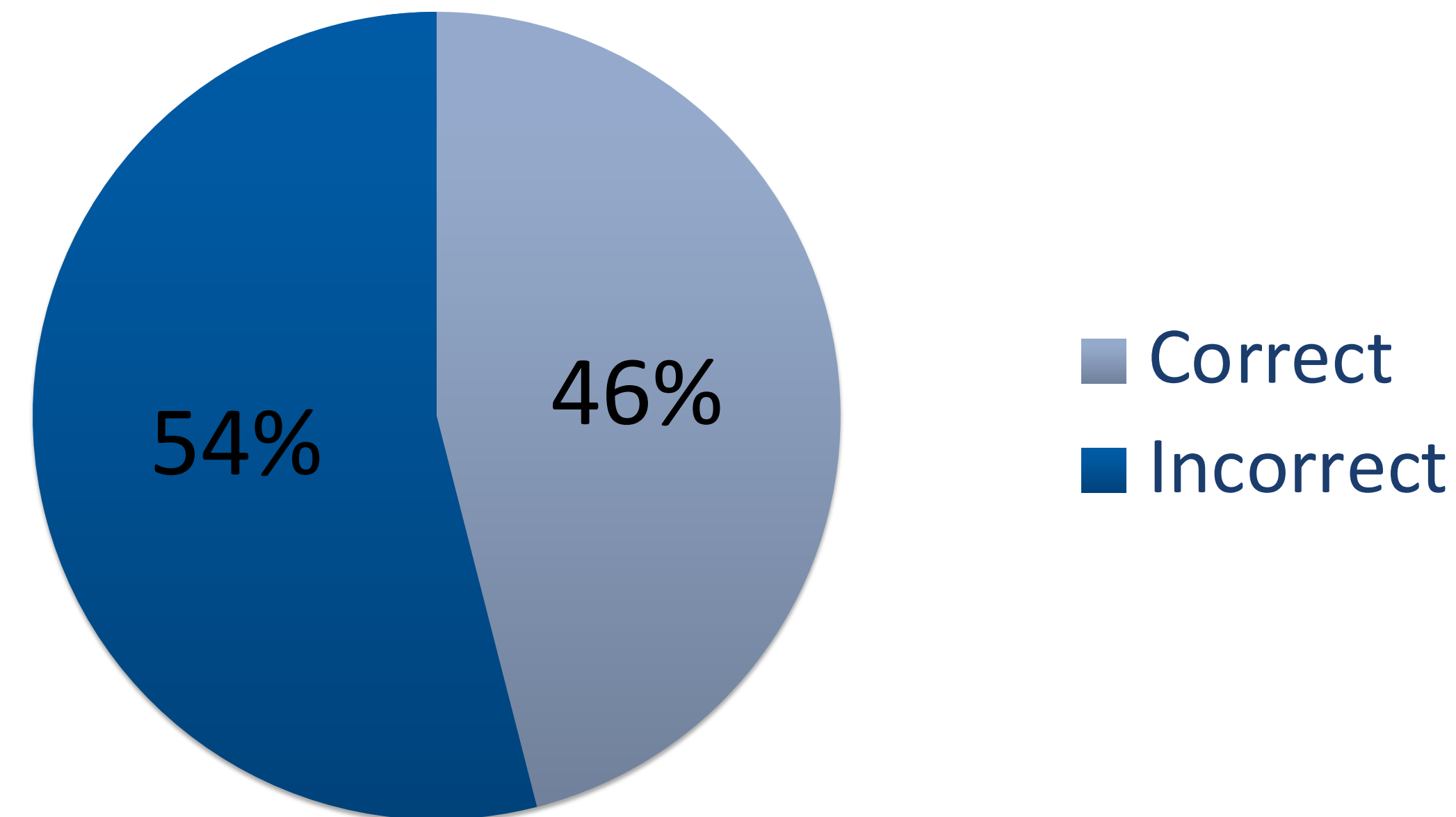
## Introduction

- According to the Pew Internet and American Life Project, 72% of American Internet users search the Internet for health information<sup>1</sup>.
- However, searchers would also encounter difficulty with expressing their desired search in a search engine: symptoms could be difficult to describe, and search engine results could ignore certain inputs while focusing on others<sup>2</sup>.
- ChatGPT (OpenAI, San Francisco, CA) is a new large language model launched by OpenAI in November 2022 programmed with reinforcement learning techniques.
- This study is the first to evaluate ChatGPT's ability to diagnose otolaryngology pathologies based on symptoms given from a patient perspective. Its public availability allows patients to input their diagnostic queries as easily as into a search engine. The conversational nature of a chatbot also may facilitate better expression of symptoms, resulting in more accurate informational outputs.

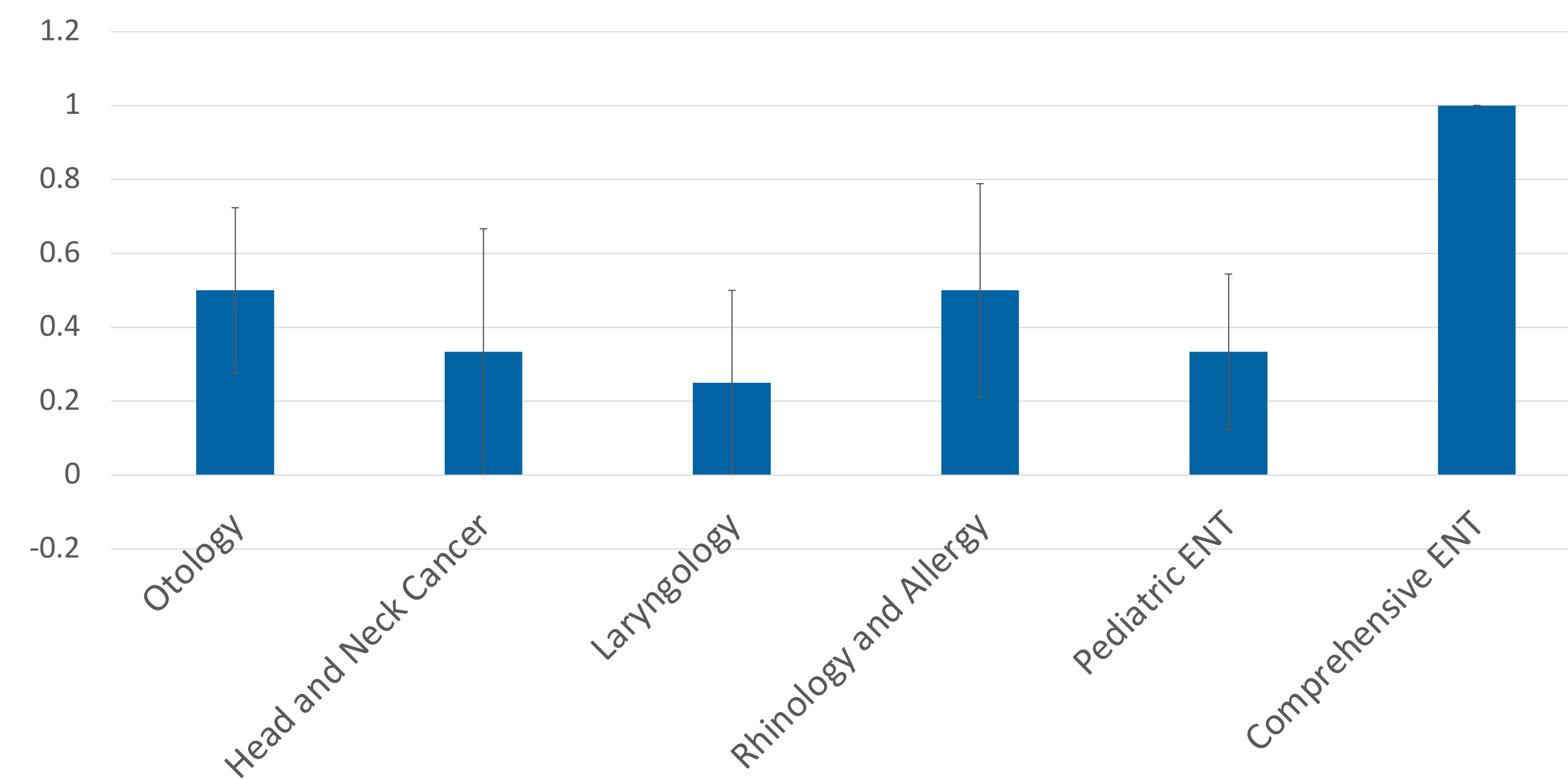
## References

1. Fox S, Duggan M. Health Online 2013. 2013. [https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP\\_HealthOnline.pdf](https://www.pewinternet.org/wp-content/uploads/sites/9/media/Files/Reports/PIP_HealthOnline.pdf) (Archived by WebCite® at <https://www.webcitation.org/74sbsnhYG>).
2. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. The Internet: A window on adolescent health literacy. *J Adolesc Heal.* 2005;37(3):243.e1-243.e7. doi:10.1016/j.jadohealth.2004.08.023

## Results



**Table 1: ChatGPT Raw Performance**  
ChatGPT's performance on including the correct diagnosis in 26 prompts.



**Table 2: Diagnosis Correct by Category**  
When further stratified by subspecialty, ChatGPT determined the correct diagnosis 40-60% of the time: Otolaryngology (50%, 3/6), Head and Neck Cancer (33.33%, 1/3), Laryngology (25%, 1/4), Rhinology and Allergy (50%, 2/4), Pediatric ENT (33.33%, 2/6), and Comprehensive ENT (100%, 3/3). In a one-way ANOVA analysis, the differences in the results across the different stratifications was not significant ( $p = 0.4865$ ).

## 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.
- 26 common otolaryngologic disorders were analyzed in this study.
- These disorders were categorized by subspecialty: Otolaryngology, Head and Neck Cancer, Laryngology, Rhinology and Allergy, Pediatric ENT, and Comprehensive ENT.
- 5 symptoms were then input into ChatGPT from a patient-centered statement.
- The results were collated and recorded in a database, which was then analyzed to determine if the AI was able to accurately diagnose the intended pathology

## Conclusion

- ChatGPT performed at <50% accuracy on including the correct diagnosis in response to a patient-perspective input of symptoms.
- With every diagnosis suggested, ChatGPT encouraged seeking a doctor's opinion.
- This study was limited by the number of inputs
- Further studies are needed to determine ChatGPT's role in first line diagnostics and aiding patients in evaluating the need for further medical advice.