The Laryngoscope

Letter to the Editor

In Reference to *The Comparative Diagnostic Capability of Large Language Models in Otolaryngology*

Dear Editor,

We read the article of Warrier et al. entitled "The Comparative Diagnostic Capability of Large Language Models in Otolaryngology.[1]" The study addresses a significant necessity in our domain as we contend with the swift incorporation of artificial intelligence (AI) into clinical practice. The authors' rigorous methodology in examining ChatGPT-3.5, Google Bard, and Bing-GPT4 using 100 clinical vignettes establishes a solid framework for evaluating these tools. Their discovery that ChatGPT-3.5 surpassed its peers with a 95.7% accuracy rate (excluding instances for further testing) is both remarkable and stimulating. We congratulate the authors for their timely and interesting study comparing the diagnostic capacities of large language models (LLMs) in otolaryngology.[1] This work well illustrates the potential of AI in otolaryngology. Some recent works demonstrated several potential applications of ChatGPT-4 in the ENT field, in particular finding LLM's reliability in analyzing laryngeal pictures.[2,3] These papers jointly highlight the increasing significance of AI in clinical decision assistance and teaching. Nevertheless, it is essential to regard these findings with tempered hope. Kleebayoon and Wiwanitkit underscored the necessity for careful incorporation of ChatGPT in clinical otolaryngology,[4] a viewpoint reiterated by Tessler et al.[5] in their assessment of ChatGPT's compliance with clinical practice guidelines. The performance diversity among various LLMs, as emphasized by Warrier et al., accentuates the necessity for careful assessment and oversight in their clinical utilization. A limitation of the study is its emphasis on diagnosis accuracy, neglecting the intricacies of clinical reasoning and the possibility of AI augmenting, rather than supplanting, human competence. Subsequent study may gain from integrating measures that evaluate the quality and pertinence of AI-generated explanations, as examined by Zalzal et al. in their assessment of ChatGPT's capacity to respond to patient inquiries.[6] Furthermore, the swiftly advancing characteristics of AI technology present a problem for comparative analyses. We would underscore the essential requirement for standardized evaluation instruments in measuring AI efficacy in medical applications. The Artificial Intelligence Performance Instrument (AIPI)[7] represents an important advancement in this area. Integrating validated instruments in future studies would improve the comparability and reliability of findings across various AI platforms and medical specializations. In the final analysis, Warrier et al. have offered significant insights on the present capabilities and constraints of LLMs in otolaryngology diagnoses. As we advance the integration of AI in our domain, research such as this will be crucial in directing responsible application and pinpointing areas for further enhancement.

Sincerely,
Antonino

Antonino Maniaci, MD, PhD
Department of Medicine and Surgery, University of Enna Kore, Enna, 94100, Italy
ASP Ragusa-Hospital Giovanni Paolo II, Ragusa, 97100, Italy
Yo-IFOS, Research Study Group, Young Otolaryngologists-International Federation of Otorhinolaryngological Societies, Paris, 13005, France

Mario Lentini, MD
Department of Medicine and Surgery, University of Enna Kore, Enna, 94100, Italy
ASP Ragusa-Hospital Giovanni Paolo II, Ragusa, 97100, Italy

Paolo Boscolo-Rizzo, MD, PhD
Department of Medical, Surgical and Health Sciences, Section of Otolaryngology, University of Trieste, Trieste, Italy

Jerome R. Lechien, MD, PhD
Yo-IFOS, Research Study Group, Young Otolaryngologists-International Federation of Otorhinolaryngological Societies, Paris, 13005, France
Department of Human Anatomy and Experimental Oncology, Faculty of Medicine, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium

Editor's Note: This Manuscript was accepted for publication on October 03, 2024.
The authors have no funding, financial relationships, or conflicts of interest to disclose.

BIBLIOGRAPHY

1. Warrier A, Singh R, Haleem A, Zaki H, Eloy JA. The comparative diagnostic capability of large language models in otolaryngology. *Laryngoscope*. 2024;134:3997-4002.
2. Maniaci A, Chiesa-Estomba CM, Lechien JR. ChatGPT-4 consistency in interpreting laryngeal clinical images of common lesions and disorders. *Otolaryngol Head Neck Surg*. 2024;171:1106-1113. https://doi.org/10.1002/ohn.897.

Send correspondence to Antonino Maniaci, Department of Medicine and Surgery, University of Enna Kore, 94100 Enna, Italy. Email: tnmaniaci29@gmail.com

DOI: 10.1002/lary.31956

3. Lechien JR, Rameau A. Applications of ChatGPT in otolaryngology-head neck surgery: a state of the art review. *Otolaryngol Head Neck Surg*. 2024;171:667-677.
4. Kleebayoon A, Wiwanitkit V. Role of ChatGPT in clinical otolaryngology. *Am J Otolaryngol*. 2024;45:104042.
5. Tessler I, Wolfovitz A, Alon EE, et al. ChatGPT's adherence to otolaryngology clinical practice guidelines. *Eur Arch Otorhinolaryngol*. 2024;281:3829-3834.
6. Zalzal HG, Abraham A, Cheng J, Shah RK. Can ChatGPT help patients answer their otolaryngology questions? *Laryngoscope Investig Otolaryngol*. 2023;9:e1193.
7. Lechien JR, Maniaci A, Gengler I, Hans S, Chiesa-Estomba CM, Vaira LA. Validity and reliability of an instrument evaluating the performance of intelligent chatbot: the artificial intelligence performance instrument (AIPI). *Eur Arch Otorhinolaryngol*. 2024;281:2063-2079.