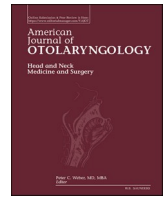Contents lists available at ScienceDirect

American Journal of Otolaryngology–Head and Neck Medicine and Surgery

journal homepage: www.elsevier.com/locate/amjoto

Evaluating AI in patient education: The need for a validated performance assessment tool

A R T I C L E   I N F O

Dear Editor,

We wish to provide a thoughtful evaluation of the recent article entitled "Answering head and neck cancer questions: An assessment of ChatGPT responses" authored by Wei *et al* [1]. The research undertaken is commendable for its examination of artificial intelligence (AI) within the sphere of patient education—a topic of increasing importance as AI tools become more prevalent among the general populace. The authors' choice to compare ChatGPT's responses with those from Google searches on the crucial subject of head and neck cancer is particularly apt and necessary [2]. The utilization of the modified Ensuring Quality Information for Patients (EQIP) tool to assess content quality lends a robustness to the study, contributing to the integrity of its conclusions [3]. Furthermore, the application of the Flesch reading ease scale to gauge the readability of information is a valuable addition, as it provides an indication of how accessible the information may be to patients [4]. While the methodology of the study is comprehensive, there are limitations that warrant discussion. The innovative use of the « People Also Ask » feature to select questions is a strength; however, it may not capture the breadth of patient inquiries [5]. A wider array of questions could provide a more all-encompassing assessment. Additionally, the inherent subjectivity of Likert scale assessments, even when standardized, introduces the potential for bias which may affect the study's outcomes [6]. A more detailed account of inter-rater reliability measures would serve to fortify the credibility of the study's findings [7]. The reliance on readability scores as a stand-in for content quality may not fully capture the clinical applicability and preciseness of the information provided. While informative, these metrics alone do not suffice to evaluate the complex nature of medical information's comprehensiveness. An oversight in the article is the failure to consider the evolutionary potential of AI models like ChatGPT (as depicted in Fig. 1). This gap in analysis neglects the iterative improvement of AI through continuous updates and learning [8]. The study also identifies a notable absence in current assessments of AI efficacy in medical information provision—an absence that could be filled by the AIPI tool, specifically developed to systematically gauge the effectiveness of AI-generated medical responses [9]. The AIPI allows to examine accuracy, completeness, and clinical pertinence, as well as user satisfaction and adherence to evidence-based practices [10]. In essence, while the article offers valuable perspectives on the capabilities and limitations of AI in patient education, the formulation and adoption of a validated assessment tool like the AIPI is imperative for future research [11,12]. Such an instrument would greatly enrich our comprehension and confidence in AI as a dependable resource in the healthcare domain [13].
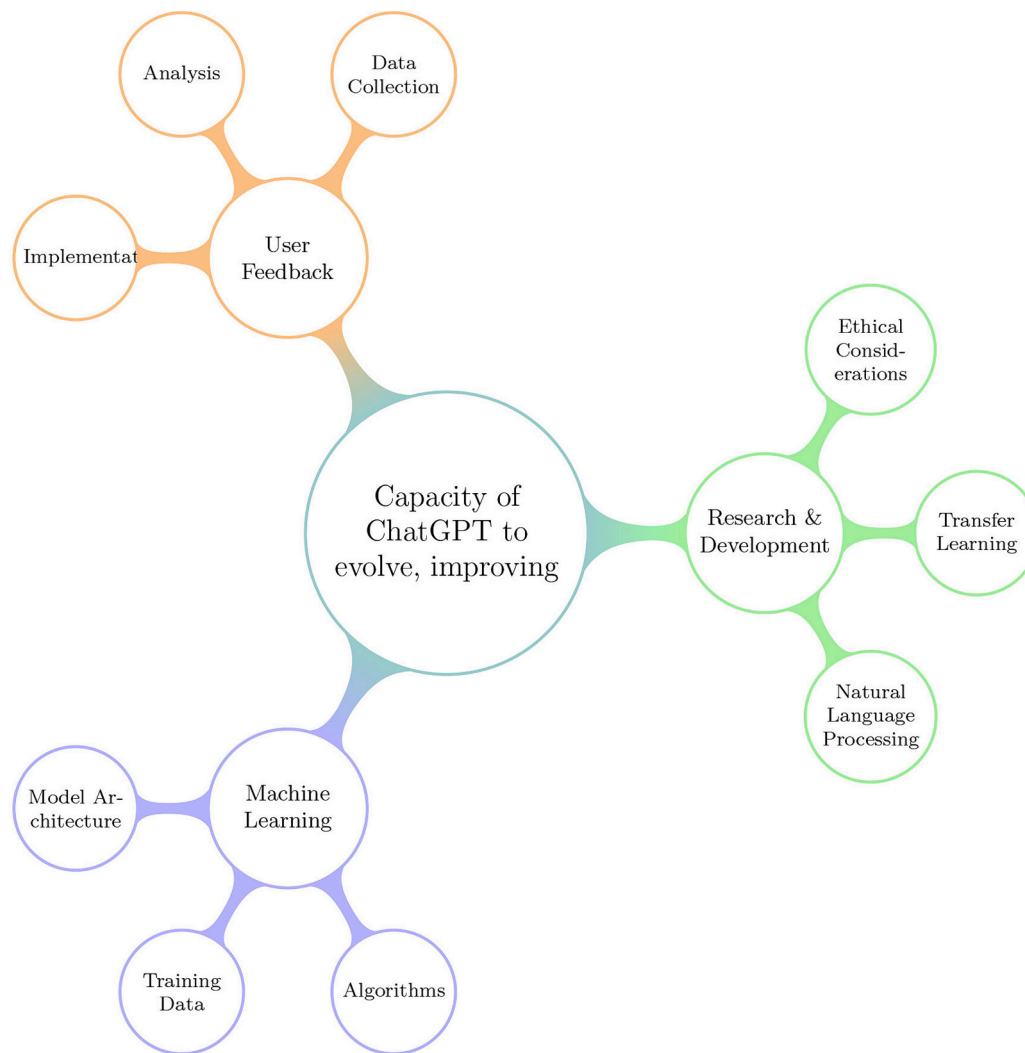
Fig. 1. Flow diagram on ChatGPT potential applications in healthcare.

CRediT authorship contribution statement

Salvatore Lavalle: Writing – review & editing. **Jerome R. Lechien:** Writing – review & editing. **Carlos Chiesa-Estomba:** Writing – review & editing. **Federica Maria Parisi:** Writing – review & editing. **Antonino Maniaci:** Writing – review & editing.

References

[1] Wei K, Fritz C, Rajasekaran K. Answering head and neck cancer questions: an assessment of ChatGPT responses. Am J Otolaryngol 2024;45(1):104085. https://doi.org/10.1016/j.amjoto.2023.104085.

[2] Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. Diagn Interv Imaging 2023;104(6):269–74. https://doi.org/10.1016/j.diii.2023.02.003.

[3] Duran GS, Yurdakurban E, Topsakal KG. The quality of CLP-related information for patients provided by ChatGPT. Cleft Palate Craniofac J 2023 Dec;21: 10556656231222387. https://doi.org/10.1177/10556656231222387 [Epub ahead of print. PMID: 38128909].

[4] Kirchner GJ, Kim RY, Weddle JB, Bible JE. Can Artificial Intelligence Improve the Readability of Patient Education Materials? Clin Orthop Relat Res 2023 Nov 1;481 (11):2260–2267. doi:https://doi.org/10.1097/CORR.0000000000002668. [Epub 2023 Apr 28. PMID: 37116006; PMCID: PMC10566892].

[5] Kasthuri V, Homer A, Alsoof D, Hong J, McDonald CL, Diebo BG, et al. Modern internet search analytics and spine: what are patients asking and reading online? N Am Spine Soc J 2023 Apr;6(14):100214. https://doi.org/10.1016/j.xnsj.2023.100214. PMID: 37214263; PMCID: PMC10192655.

[6] Westland JC. Information loss and bias in likert survey responses. PloS One 2022 Jul 28;17(7):e0271949. https://doi.org/10.1371/journal.pone.0271949. PMID: 35901102; PMCID: PMC9333316.

[7] Currie G, Singh C, Nelson T, Nabasenja C, Al-Hayek Y, Spuur K. ChatGPT in medical imaging higher education. Radiography (Lond) 2023;29(4):792–9. https://doi.org/10.1016/j.radi.2023.05.011.

[8] Manco L, Maffei N, Strolin S, Vichi S, Bottazzi L, Strigari L. Basic of machine learning and deep learning in imaging for medical physicists. Phys Med 2021;83: 194–205. https://doi.org/10.1016/j.ejmp.2021.03.026.

[9] 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 2023. https://doi.org/10.1007/s00405-023-08219-y.

[10] Lechien JR, Chiesa-Estomba CM, Baudouin R, Hans S. Accuracy of ChatGPT in head and neck oncological board decisions: preliminary findings. Eur Arch Otorhinolaryngol 2023. https://doi.org/10.1007/s00405-023-08326-w.

[11] Streiner DL, Saboury B, Zukotynski KA. Evidence-based artificial intelligence in medical imaging. PET Clin 2022;17(1):51–5. https://doi.org/10.1016/j.cpet.2021.09.005.

[12] Srivastav S, Chandrakar R, Gupta S, Babhulkar V, Agrawal S, Jaiswal A, et al. ChatGPT in radiology: the advantages and limitations of artificial intelligence for medical imaging diagnosis. Cureus 2023;15(7):e41435. https://doi.org/10.7759/cureus.41435.

[13] Varney ET, Lee CI. The potential for using ChatGPT to improve imaging appropriateness. J Am Coll Radiol 2023;20(10):988–9. https://doi.org/10.1016/j.jacr.2023.06.005.

Salvatore Lavalle[a], Jerome R. Lechien[b,c], Carlos Chiesa-Estomba[b,d], Federica Maria Parisi[e], Antonino Maniaci[a,b,*]

[a] *Faculty of Medicine and Surgery, University of Enna Kore, 94100 Enna, Italy*

[b] *Head & Neck Study Group, Young-Otolaryngologists of the International Federations of Oto-Rhino-Laryngological Societies (YO-IFOS), 13005 Marseille, France*

[c] *Department of Human Anatomy and Experimental Oncology, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium*

[d] *Department of Otorhinolaryngology - Head and Neck Surgery, Donostia University Hospital, San Sebastian, Spain*

[e] *Department of Medical and Surgical Sciences and Advanced Technologies "GF Ingrassia", ENT Section, University of Catania, Via S. Sofia, 78, 95125 Catania, Italy*

[*] Corresponding author at: Faculty of Medicine and Surgery, University of Enna Kore, 94100 Enna, Italy.

E-mail addresses: salvatore.lavalle@unikore.it (S. Lavalle), jerome.lechien@unimons.ac.be (J.R. Lechien), tnmaniaci29@gmail.com (A. Maniaci).