AMERICAN ACADEMY OF OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

State of the Art Review

Otolaryngology–Head and Neck Surgery
2024, Vol. 171(3) 667–677

DOI: 10.1002/ohn.807
http://otojournal.org
WILEY

Applications of ChatGPT in Otolaryngology–Head Neck Surgery: A State of the Art Review

Jérôme R. Lechien, MD, PhD, MS, AFACS[1,2,3,4], and Anais Rameau, MD, MPhill, FACS[5]

Abstract

Objective. To review the current literature on the application, accuracy, and performance of Chatbot Generative Pre-Trained Transformer (ChatGPT) in *Otolaryngology–Head and Neck Surgery*.

Data Sources. PubMED, Cochrane Library, and Scopus.

Review Methods. A comprehensive review of the literature on the applications of ChatGPT in otolaryngology was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-analyses statement.

Conclusions. ChatGPT provides imperfect patient information or general knowledge related to diseases found in *Otolaryngology–Head and Neck Surgery*. In clinical practice, despite suboptimal performance, studies reported that the model is more accurate in providing diagnoses, than in suggesting the most adequate additional examinations and treatments related to clinical vignettes or real clinical cases. ChatGPT has been used as an adjunct tool to improve scientific reports (referencing, spelling correction), to elaborate study protocols, or to take student or resident exams reporting several levels of accuracy. The stability of ChatGPT responses throughout repeated questions appeared high but many studies reported some hallucination events, particularly in providing scientific references.

Implications for Practice. To date, most applications of ChatGPT are limited in generating disease or treatment information, and in the improvement of the management of clinical cases. The lack of comparison of ChatGPT performance with other large language models is the main limitation of the current research. Its ability to analyze clinical images has not yet been investigated in otolaryngology although upper airway tract or ear images are an important step in the diagnosis of most common ear, nose, and throat conditions. This review may help otolaryngologists to conceive new applications in further research.

Keywords

artificial Intelligence, ChatGPT, generative, GPT, head neck, otolaryngology, otorhinolaryngology, surgery

Received February 20, 2024; accepted April 19, 2024.

The development of artificial intelligence (AI)-powered language models, such as Chatbot Generative Pre-Trained Transformer (ChatGPT) is emerging in medicine and surgery. ChatGPT is one of the most used chatbots, while the number of studies dedicated to the application of ChatGPT is increasing. ChatGPT has been found to respond to simple and complicated questions related to clinical and basic science research,[1] referencing,[2] medical examinations,[3] clinical vignettes,[4] and may improve scientific reports through spelling correction.[5] To date, ChatGPT is accessible by an estimated 100 million users,[6] which may encourage patients to use it for education prior to a medical consultation,[6] while some young practitioners may consider ChatGPT as an adjunct clinical tool for improving their practice.[7] Since the first report of ChatGPT application in *Otolaryngology–Head and Neck Surgery*,[8] the number of studies assessing the applications, accuracy, and performance is constantly increasing.[9-12] The aim of this state of the art review was to review the current literature on the application, accuracy, and performance of ChatGPT in *Otolaryngology–Head and Neck Surgery* at the 1-year anniversary or the release of ChatGPT to the public.

[1]Research Committee of Young Otolaryngologists of the International Federation of Otorhinolaryngological Societies (IFOS), Paris, France
[2]Division of Laryngology and Broncho-Esophagology, Department of Otolaryngology–Head Neck Surgery, EpiCURA Hospital, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium
[3]Department of Otorhinolaryngology and Head and Neck Surgery, Foch Hospital, Phonetics and Phonology Laboratory (UMR 7018 CNRS, Université Sorbonne Nouvelle/Paris 3), Paris Saclay University, Paris, France
[4]Department of Otorhinolaryngology and Head and Neck Surgery, CHU Saint-Pierre, Brussels, Belgium
[5]Department of Otolaryngology–Head and Neck Surgery, Sean Parker Institute for the Voice, Weill Cornell Medicine, New York City, New York, USA

Corresponding Author:
Jérôme R. Lechien, MD, PhD, MS, AFACS, Division of Laryngology and Broncho-Esophagology, Department of Otolaryngology–Head Neck Surgery, EpiCURA Hospital, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium.
Email: Jerome.Lechien@umons.ac.be

Methods

A PubMed, Cochrane Library, and Scopus database research was conducted for relevant peer-reviewed publications in English related to the application, accuracy, and performance of ChatGPT in *Otolaryngology–Head and Neck Surgery*. The following terms were used: "ChatGPT," "GPT," "Chatbot," "Otolaryngology," "Head Neck," "Surgery," and "Ear Nose Throat." Studies assessing ChatGPT application in the management of clinical vignettes, true clinical cases, surgery, scientific referencing, scientific paper improvement, patient information, or any other fields related to otolaryngology were considered. Case reports or letters to the editor without data were excluded. From this initial review of the literature, publications were selected for inclusion in the final review if authors provided enough information on methods used to conduct the study. Article selection by Preferred Reporting Items for Systematic Reviews and Meta-analyses criteria[13] is summarized in the flowchart in **Figure 1**. The ChatGPT was used in several fields related to *Otolaryngology–Head and Neck Surgery*, including the management of clinical vignettes, real clinical cases, the improvement of scientific publications or referencing, and generating patient or practitioner information (**Figure 1**). Critical analysis of this literature focused on the accuracy, performance, limitations, and strengths of ChatGPT in the related applications. Implications for practice were then summarized. Ethics committee approval was not required for this review.

Discussion

Patient Counseling, Guidance, and Education

The large majority of studies conducted in otolaryngology have investigated the accuracy of ChatGPT in patient counseling, guidance, and education in general otolaryngology,[1,14-19] pediatric otolaryngology,[20] otology,[21] facial and plastic surgery,[22-25] endocrine otolaryngology,[26] sleep,[27-29] head and neck oncology,[30-34] and laryngology (**Table 1**).[35] Among them, the information provided by ChatGPT was compared to those provided by browser search or other large language models (LLMs) or websites, for example, Google Bard,[29,31] Google search.[16-18,21]

Seven investigations studied the accuracy of ChatGPT in general otolaryngology information.[1,14-19] Depending on the method used to evaluate the accuracy, 56.7% to 89% of information provided by ChaGPT-3.5 or 4.0 were judged as accurate for education, medical advice, or understandability (**Table 1**).[1,14-19] The ChatGPT information matched with the current guidelines of the American Academy of Otolaryngology–Head and Neck Surgery in 68% to 89%,[16] but Google search reported overall better results than ChatGPT-3.5 in terms of readability score (Flesch Reading Ease) or understandability (Flesch-Kincaid Grade Level).[16,18] Among these studies, Vaira et al reported that ChatGPT-4 provided 50% of false references to support some answers,[1] while Langlie et al did not find major errors or hallucinations.[19] Interestingly, Zalzal et al demonstrated that the initial ChatGPT-3.5 full and partial accuracy rates (56.7% and 86.7%) significantly improved (73.3% and 96.7%) after regenerated input into the application programming interface (API) and human feedback.[14] The high accuracy of ChatGPT-3.5 in providing patient information regarding indications, procedures, alternative therapeutic options for common otolaryngological surgeries (eg, adenotonsillectomy, tympanoplasty, endoscopic sinus surgery, parotidectomy, total laryngectomy) was, however, tempered by authors who noted that ChatGPT-3.5 lacked precision in the description of procedural steps, details, and risks related to procedures.[19]

The accuracy of ChatGPT-3.5 in pediatric otolaryngology was evaluated towards parent information for tympanostomy.[20] ChatGPT-3.5 provided 95.7% of accurate responses regarding the guidelines of the American Academy of Otolaryngology–Head and Neck Surgery.[20]

In the same vein, the accuracy of the several versions of ChatGPT was high for general information related to upper aerodigestive tract cancers.[30-34] However, the performance of ChatGPT decreased when authors investigated its ability to provide information dedicated to the follow-up of oropharyngeal cancer.[30] Lee et al compared the accuracy of ChatGPT-3.5 versus simple browser search for indications, risks, and recovery time for 5 common head and neck surgeries.[32] In this study, judges preferred the ChatGPT-3.5 responses compared to those of publicly available websites in 48% of cases.[32] ChatGPT-3.5 was compared with ChatGPT-4 and Google Bard in terms of safety and global quality of information provided for laryngeal cancer.[31] Surprisingly, authors reported better results for ChatGPT-3.5 over ChatGPT-4 and Google Bard while ChatGPT-4 was expected to be better than the free version.[31] In head and neck surgery, only Kuscu et al investigated the stability of ChatGPT-4 in regenerated questions, which reached 94.1% of similar regenerated outputs.[33] In this study, ChatGPT-4 reported 100%, 92.6%, 88.9%, and 80% accurate rates for prevention, diagnosis, treatment, surgical features (recovery, risks, complications, and follow-up), respectively.[33] Among the several cancer locations, Chiesa-Estomba et al observed that patients similarly appreciated the information of practitioners and ChatGPT-3.5 for salivary gland cancer but preferred the explanations of practitioners for laryngeal and oropharyngeal cancers.[34]

The knowledge of ChatGPT was investigated in sleep medicine in 4 studies.[27-29,36] Cheong et al reported in a first study that GPT-4 successfully achieved the pass mark of 80% in 5 of the 10 exam categories of the American Academy of Sleep Medicine, which included the Normal Sleep and Variants Self-Assessment Exam (2021), Circadian Rhythm Sleep-Wake Disorders Self-Assessment Exam (2021), Insomnia Self-Assessment Exam (2022), Parasomnias Self-Assessment Exam (2022) and the Sleep-

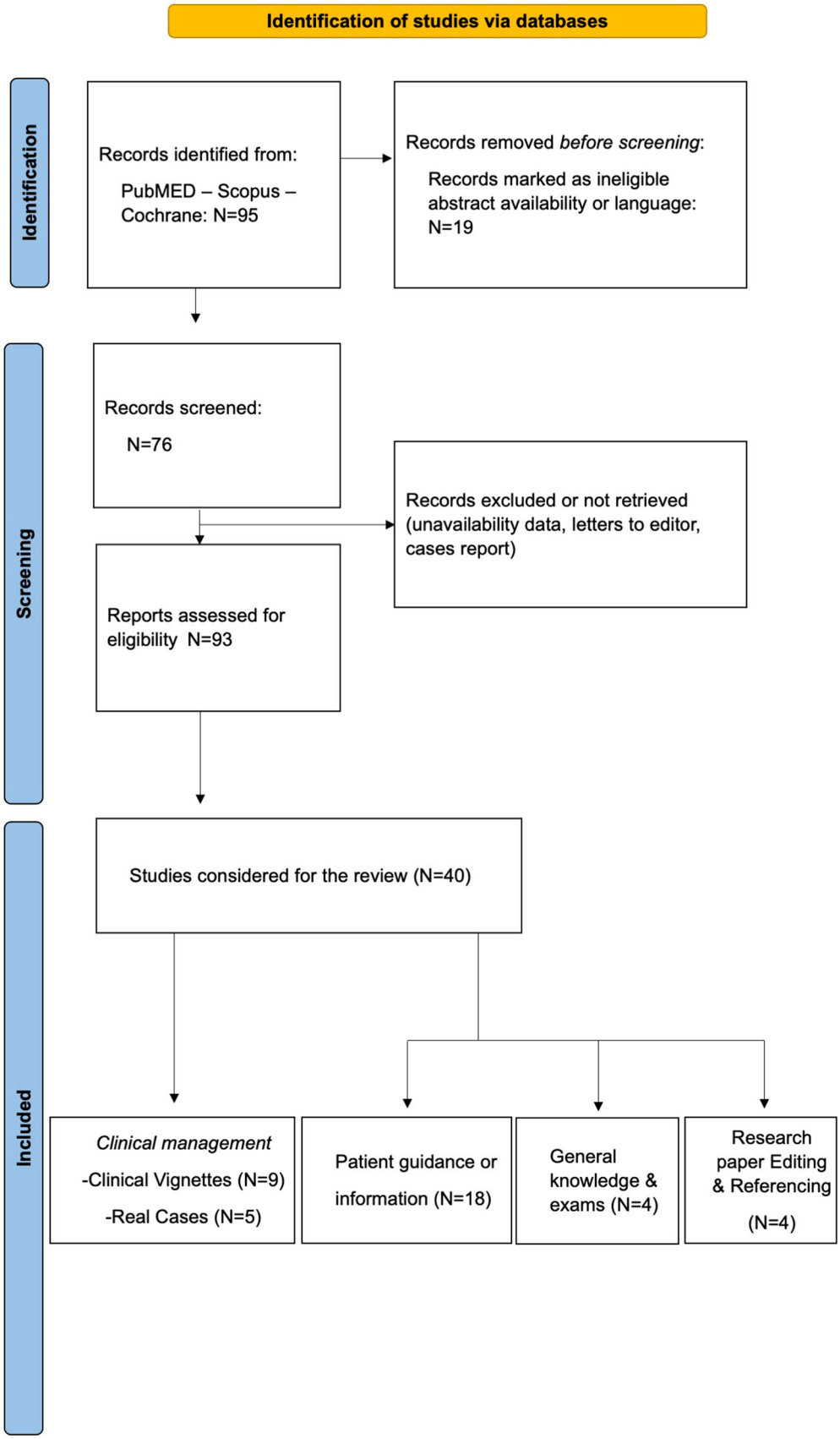

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-analyses flowchart.

Table 1. Application in Otolaryngological Knowledge and Patient Information

References	Accuracy assessment	Design	GPT	N	Tool	Outcomes	Results
Ayoub et al[16]	GPT/Google search info	Prospective-CS	uGPT	Questions: 10	PEMAT-P	Medical advice	68%-89%
	General otolaryngology		Google	AAOHNS GL	GPT vs GS	Patient education	87%-78%
Vaira et al[1]	Information and knowledge	Prospective-CS	GPT-4	Questions: 144	6-point	open question—AC-CO	5.43-2.7
	General otolaryngology			Experts: 18	Likert scale	Closed question (AC, %)	84.7%
Shen et al[18]	GPT/Google search info	Prospective-CS	GPT-3.5	Questions: 54	PEMAT	Understandable (PEMAT)	Google=GPT
	General otolaryngology		Google	AAOHNS GL	FKGL-FRE	FKGL-FRE	Google>GPT
Langlie et al[19]	Patient information	Prospective-CS	GPT-3.5	Questions: 5	None	Major aberrant errors	0
	Otolaryngology surgeries			Experts: 2 PO		Appropriate treatment	5/5
Nielsen et al[15]	Patient information	Prospective-CS	GPT-4.0	Questions: 27	5-point	Mean score (AC, RE, DE)	3.51-3.71-3.00
	General otolaryngology			Experts: 13	Likert scale		
Zalzal et al[14]	Clinical questions	Prospective-CS	GPT-3.5	Questions: 30	3-point	Accuracy GPT	56.7%-73.3%
	General otolaryngology			Experts: 2 PO	scale	Partial accuracy	86.7%-96.7%
Moise et al[20]	Patient information	Prospective-CS	GPT-3.5	Questions: 23	Binary	Accurate responses	22/23 (95.7%)
	Pediatric tympanostomy			AAOHNS GL	(correct/no)		
Lee et al[32]	Preoperative counseling	Prospective-CS	GPT-3.5	Surgeries: 5	5-point	Accuracy GPT	5/5
	Head and neck surgery		BR	Experts: 5 PO	Likert scale	Expert preference	48% for GPT info
Davis et al[30]	Patient information	Prospective-CS	GPT-3.5	Questions: 15	5-point	AC, CA, SA	3.88-3.80-3.67
	Oropharyngeal cancer			Experts: 4 PO	Likert scale	AC	Posttreatment>diag.
Chiesa-Estomba et al[34]	Patient opinion in GPT head	Prospective-CS	GPT-3.5	Questions: 5	10-point	Larynx-oropharynx	Experts>GPT
	neck oncological advice			Experts: 10	Likert scale	Salivary gland	Experts=GPT
Kuscu et al[33]	Patient information	Prospective-CS	GPT-4.0	Questions: 154	4-point	Total/partial/incorrect	86.4%-11%-2.6%
	Head and neck oncology			Experts: 2 PO	Likert scale	Stability of responses	94.1%
Ostrowska et al[31]	Patient information	Prospective-CS	GPT-4/3.5	Questions: 36	3- and 5-point	Safety score	GPT-3.5>GPT-4>GB
	Laryngeal cancer		Google	Experts: 9	Likert scale	Global quality score	GPT-3.5>GPT-4>GB
Lechien et al[35]	Patient information	Prospective-CS	GPT-4.0	Questions: 25	5-point	EP, CP, AE, T AC	4.1-4.5-3.8-4.2
	Laryngopharyngeal reflux			Experts: 4 PO	Likert scale		
Campbell et al[28]	Patient information	Prospective-CS	GPT-3.5	Questions: 24	4-item	Correct/perfect	58%-14%
	Sleep disorders			Experts: 2 RO	Likert scale	Incorrect/partially correct	1%-27%
Mira et al[27]	Patient information	Prospective-CS	GPT-3.5	Questions: 10	5-point	Common responses	75% of OTO/GPT
	Sleep apnea			Experts: 97	Likert scale	Performance	Experts>GPT
Cheong et al[29]	GPT/Google bard information	Prospective-CS	GPT-3.5	Questions: 46	PEMAT-P	Understandability	GPT>Google Bard
	in sleep disorder		Google	Experts: 2		Actionability	GPT>Google Bard
Bellinger et al[21]	GPT/Google search info	Prospective-CS	GPT-3	Questions: 29	PEMAT-P	FKGL-FRE	Google>GPT
	Benign vertigo		Google	Experts: 3	FKGL-FRE	GPT AC (5-likert scale)	4.2
Campbell et al[26]	Patient information	Prospective-CS	GPT-3.5	Questions: 30	4-item	Correct/correct and referenced	47%-22%
	Thyroid and endocrine			Experts: 2 RO	Likert scale	Incorrect/partially correct	2%-28%

Abbreviations: AAOHNS GL, American Academy of Otolaryngology Guidelines; AC, average accuracy; AE, additional examination; BR, browser search; CA, comprehensive accuracy; CO, completeness; CP, clinical presentation; CS, cross-sectional; DE, depth; diag., diagnosis; EP, epidemiology; FKGL, Flesch-Kincaid Grade Level; FRE, Flesch Reading Ease; GPT, Generative Pre-Trained Transformer; GS, Google Search; N, number; OTO, otolaryngologists; PEMAT-P, Patient Education Material Assessment Tool; PO, practitioner in otolaryngology; RE, relevance; RO, resident otolaryngologists; SA, similarity accuracy; uGPT, unspecified version of GPT.

Related Movements Self-Assessment Exam (2023).[36] Authors also reported better overall results of ChatGPT-4 over ChatGPT-3.5 (68.1% vs 46.8%) and Google Bard (45.5%), while ChatGPT-3.5 and Google Bard did not report significant differences.[36] A second study of this team reported that ChatGPT-3.5 scored higher than Google Bard in providing patient information for sleep disorders (understandability and actionability).[29] Mira et al interrogated 97 sleep practitioners and ChatGPT-3.5 for providing patient information related to sleep conditions and reported 75% of common answers despite higher performance of experts in term of overall accuracy.[27] The accuracy rates of ChatGPT-3.5 were however lower in the study of Campbell et al who reported only 14% and 58% of perfect and correct responses to sleep disorder questions.[28] The same team has conducted the only study investigating the accuracy of ChatGPT-3.5 in patient information on thyroid nodules and diseases.[26] Using a 4-item Likert scale assessment, they reported that ChatGPT-3.5 provided 47% and 22% of correct, and correct/referenced responses,[26] which is slightly worse than the findings of studies conducted in general otolaryngology or head and neck oncology (**Table 1**).

The performance of ChatGPT-3 was additionally compared to Google search information for benign vertigo.[21] In this study using the third version of ChatGPT, authors reported better results of Google compared to ChatGPT-3 in terms of readability (Flesch Reading Ease) and understandability (Flesch-Kincaid Grade Level).[21]

In laryngology, ChatGPT-4 was interrogated for general knowledge on laryngopharyngeal reflux, and the responses were confronted to the Dubai Consensus on Laryngopharyngeal Reflux.[35,37] Then, 4 laryngologists reported that the ChatGPT-4 accuracy was higher for clinical presentation and treatment compared to epidemiology and additional examinations. Precisely, ChatGPT-4 was not up to date regarding the use of impedance-pH monitoring and the existence of weakly acid or alkaline laryngopharyngeal reflux.[35] The worst performance by ChatGPT was noted in providing up-to-date information in additional testing of clinical vignettes or true clinical cases.[4,9] ChatGPT's lack of consideration of the latest guidelines led Workman et al to develop a LangChain/OpenAI API-powered chatbot including International Consensus Statement of Allergy and Rhinology Rhinosinusitis (ICAR-RS), which is able to provide users with latest information and recommendations in this field.[38] The works of Workman et al is particularly relevant regarding the limitation of ChatGPT and other similar platforms for providing accurate and evidence-based information. Most LLMs source their information from publicly available websites or platforms, such as Bing, while the most reliable and evidence-based resources are often found in textbooks, and clinical sources (UpToDate) that are not open access. This limitation has been insufficiently investigated in the current literature and the inaccurate information provided by ChatGPT and other LLMs could lead to misinformation for patients. The development of LLMs dedicated to medical information based on accurate database information is of major importance to improve patient information. In trying to address the inaccuracies and hallucinations produced by ChatGPT, Workman et al[38] have trained the OpenAI GPT-3.5 API on the ICAR-RS document to improve accuracy and reliability of LLM responses and align them with evidence-based medical information. While their pilot Chatbot did not perform significantly better than ChatGPT-3.5, Workman et al aim to continue refining their approach towards improved LLM query output.

Otolaryngological General Knowledge and Exams

The first ChatGPT study published in *Otolaryngology–Head and Neck Surgery* was conducted by Hoch et al, who evaluated the accuracy of ChatGPT-3 response to 2576 practice quiz questions designed for German otolaryngology board certification.[3] Authors reported an overall accurate rate of 57% and they observed that ChatGPT-3 better responded to single-choice questions compared to multiple-choice questions (34% vs 63%). This study suggested that the performance of ChatGPT-3 was particularly high in allergology (72%) and low in the legal field (30%).[3] The better performance of ChatGPT in single-choice questions was similarly supported in another study where authors input 1800 test questions from the American, Italian, French, United Kingdom, Indian, and Spanish medical licensing examination into the API of ChatGPT-3.5.[10] They reported 22% to 73% of correct answers; the highest success being for the Italian exam, while the French exam reported the lowest performance. As observed by Hoch et al,[3] ChatGPT-3.5 reported greatest difficulties for questions exhibiting multiple correct responses.[10] Mahajan et al investigated the performance of ChatGPT-3.5 in responding to practice exam questions in *Otolaryngology–Head and Neck Surgery*. From a methodological standpoint, authors compared the ChatGPT outputs with the benchmark of the answers and explanations and they reported that ChatGPT-3.5 correctly answered 53% of questions, and provided correct explanations in 54% of cases, respectively.[11] The performance of ChatGPT-4 was investigated in only 1 study.[12] Thus, 21 common questions of licensing exam in otolaryngology were submitted to ChatGPT-4 and the related outputs were analyzed by 2 independent practitioners with the Concordance, Validity, Safety, Competency model.[12] ChatGPT-4 scored 23.5/34 (accurate rate: 69.1%) but did not reach the minimum passing score for the examination (70%). Interestingly, authors provided further queries with explicitly indicating the focus of otolaryngology field, which led ChatGPT-4 to improve

its score, reaching an accurate rate of 75%.[12] In summary, the studies supported high but not fully accurate results with passing grade across different exam formats.

Scientific Publication Improvement or Referencing

ChatGPT was initially suggested as an interesting adjunct tool to enhance the quality and efficiency of scientific manuscript writing, acting as a reliable tool against linguistic imperfections.[39-41] However, this was not confirmed in a recent study investigating the ability of ChatGPT for editing a manuscript in *Otolaryngology–Head and Neck Surgery*.[5] In this study, authors submitted 4 manuscripts written by a French native speaker, and 2 English native linguists compared the ChatGPT editing with that of a professional editing service.[5] The linguists observed that ChatGPT-4 detected 86/171 errors (50.3%), and proposed appropriate corrections for 83.7% of detected errors. Interestingly, ChatGPT-4 claimed to change something that was already there in 82 cases, which may be considered as AI-related hallucinations.[5]

ChatGPT has been investigated for searching scientific database references in 2 studies.[2,42] In the first one, Frosolini et al presented 20 clinical questions across different head and neck disciplines to ChatGPT-3.5 and ChatGPT-4.0 to produce texts on the assigned topics.[42] Based on the ChatGPT outputs, authors observed that both versions displayed a tendency to provide erroneous references to support the content of the responses (hallucinations) even if ChatGPT-4.0 outperformed version 3.5 in terms of reference reliability.[42] The risk of hallucinations and misreferences was supported in another study assessing the ability and accuracy of ChatGPT-3.5 and ChatGPT-4 to reference key papers in *Otolaryngology–Head and Neck Surgery*, which consisted of the top-30 most cited papers in the past 40 years.[2] The accuracy of ChatGPT-4 ranged from 73% to 87%, while the accuracy of ChatGPT-3.5 was significantly lower, ranging from 47% to 60%, respectively. ChatGPT-3.5 and ChatGPT-4 provided 19 and 13 inaccurate references, respectively. Three references were invented by the chatbot (2 by the ChatGPT-3.5 and 1 by the ChatGPT-4, respectively). This study also demonstrated a poor reliability of ChatGPT-3.5 and ChatGPT-4 for regenerated answers.[2] UpToDate appears to be a better tool than ChatGPT in providing scientific references according to Karimov et al who observed that UpToDate supported all clinical outputs with references from peer-reviewed journals, conference papers, or book chapters, whereas ChatGPT-3.5 did not give references in some questions.[43]

Clinical Vignette Assessment

Nine studies have investigated the accuracy and performance of ChatGPT in the management of theoretical clinical vignettes (**Table 2**).[1,4,43-48] In studies using a 5-point Likert scale, the mean score of ChatGPT ranged from 2.89 to 4.4, varying according to specialties.[4,47,48] In the field of head and neck surgery, Chiesa-Estomba et al confronted 10 practitioners versus ChatGPT-3.5 in the management of salivary gland diseases. In this study, authors reported that 10 practitioners scored higher than ChatGPT-3.5 in the overall management of 6 clinical vignettes dedicated to salivary gland conditions (4.1 vs 3.4) but ChatGPT-3.5 and practitioners similarly scored in providing information related to the treatment scores (3.3 vs 2.6).[4] Based on the National Comprehensive Cancer Network Guidelines, Marchi et al reported ChatGPT-4 accuracy rates of 85% and 96% in providing primary and alternative treatments for 272 head and neck oncological vignettes.[49] The 5-point Likert scale scores of ChatGPT-3.5 were moderate in terms of consistency (2.89) and correctness (3.80) scores in the study of Dallari et al who observed that the performance of ChatGPT-3.5 did not vary regarding the difficulty of the clinical case.[47]

Teixeira et al reported higher scores of ChatGPT-3.5 compared to others[4,47] in the management of clinical vignettes in general otolaryngology.[48] The high performance of ChatGPT (unspecified version) was similarly observed in providing recommendations for the management of dizziness.[46] The overall high accuracy scores of ChatGPT in the management of clinical vignettes[4,45,46,48] were tempered by Saibene et al[45] who reported mild-to-moderate scores of both ChatGPT-3.5 and ChatGPT-4 in providing recommendations in clinical vignettes of odontogenic rhinosinusitis,[45] and Vaira et al who reported only 56.7% of correct treatments in general otolaryngology and maxillofacial surgery practices.[1]

Clinical Decision Support Tool in Real Clinical Scenarios

Five studies were conducted to evaluate the performance of ChatGPT as an adjunct tool for the management of real clinical cases in general otolaryngology, head and neck oncology, laryngology, or rhinology (**Table 3**).[50-54] The same team used the Artificial Intelligence Performance Instrument for the assessment of the ability of ChatGPT-4 to provide correct primary and differential diagnoses, the most adequate additional examinations and treatments.[50,52,53] Regardless of the subspecialty, ChatGPT-4 reported highest accuracy in providing a plausible primary diagnosis (47%-79%), and lowest scores for selecting the most adequate additional examinations (**Table 3**). Indeed, authors noted that the chatbot works as a virtual encyclopedia that proposed a list of potential additional examinations without selection of the most adequate.[50,52,53] In laryngology, ChatGPT-3.5 was interrogated to make a differential diagnosis of 40 clinical cases, and to indicate the most adequate additional examinations.[51] As for the aforementioned studies,[50,52,53] the number of additional examinations recommended by ChatGPT-3.5 was significantly higher than those indicated by practitioners, while it reported highest accuracy for making a primary and differential diagnosis.[51] Interestingly, as suggested by

Table 2. Accuracy of ChatGPT in Clinical Vignettes

References	Specialty	Design	N	Features	Tool	Outcomes	Results
Chiesa-Estomba et al[4]	Salivary glands	Prospective	6 CV	GPT-3.5	5-point Likert scale	Mean agreement PO vs GPT	4.1-3.4/5 PO>GPT
		Cross-sectional	10 PO			Treatment N	2.6-3.3/5 PO=GPT
Marchi et al[49]	Head and neck oncology	Prospective	272 CV	GPT-3.5	0-1 coefficient	PT, AT (SE, AC)	100%-85%-100%-96%
		Cross-sectional	NCCNG	NCCN vs GPT		Follow-up indication SE, AC	100%-94%/1%
Qu et al[44]	General otolaryngology	Prospective	20 CV	GPT-4	5-point Likert scale	Differential diagnosis	AP=GPT
		Cross-sectional	11 PO			Treatment plan	
Karimov et al[43]	General otolaryngology	Prospective	25 CV	GPT-3.5	Composite	Usefulness composite score	UpToDate>GPT
		Cross-sectional	2 experts	UpToDate	score	CV related references	UpToDate>GPT
Saibene et al[45]	Rhinology and dental	Prospective	5 CV	GPT-3.5	Total disagreement	Total disagreement score	GPT-3.5>GPT-4
		Cross-sectional	8 experts	GPT-4	score	Disagreement expert vs GPT	73/80
Chee et al[46]	Dizziness	Prospective	8 CV	uGPT	Not provided	Accurate diagnosis	6/8 cases
		Cross-sectional	NP				
Dallari et al[47]	General otolaryngology	Prospective	20 CV	GPT-3.5	5-point Likert scale	Correctness score	3.80
		Cross-sectional	3 experts			Consistency score	2.89
Teixeria et al[48]	General otolaryngology	Prospective	20 CV	GPT-3.5	5-point Likert scale	Mean score (GPT vs PO)	4.4-4.92
		Cross-sectional	5 PO			Regenerated mean score	4.15
Vaira et al[1]	General otolaryngology	Prospective	15 CV	GPT-4	None	Nearly or fully correct PD	81.7%
	Maxillofacial surgery	Cross-sectional	18 experts			Correct treatment	56.7%

Abbreviations: AC, accuracy; AP, attending physicians; AT, adjuvant treatment; CV, clinical vignettes; GPT, Generative Pre-Trained Transformer; NCCNG, National Comprehensive Cancer Network Guidelines; NP, not provided; PD, primary diagnosis; PO, practitioner in otolaryngology; SE, sensitivity; uGPT, unspecified version of GPT.

Table 3. Accuracy of ChatGPT in the Management of Clinical Cases

Authors	Specialty	Design	N	Features	Tool	Outcomes	Results
Lechien et al[49]	Head and neck oncology	Prospective	20 CC	GPT-4	AIPI	AE prescription (N)	GPT>OTO
		Cross-sectional	2 PO	PO vs GPT		AE, TT, cTNM	25%-55%-95%
Lechien et al[50]	Laryngology	Prospective	40 CC	GPT-3.5	OCAT	AE prescription (N)	GPT>OTO
		Cross-sectional	2-3 PO	PO vs GPT		AE, TT, PD/DD	10%-33%-60%-79%-65%
Radulesco et al[51]	Rhinology and allergy	Prospective	40 CC	GPT-4	AIPI	AE prescription (N)	GPT>OTO
		Cross-sectional	3 PO	PO vs GPT		AE, TT, PD	7%-30%-7%-32%-47%-55%
Lechien et al[52]	General otolaryngology	Prospective	45 CC	GPT-4	AIPI	AE prescription (N)	GPT>OTO
		Cross-sectional	2 PO	PO vs GPT		AE, TT, PD	29%-22%-56%-71%
Sievert et al[53]	Head and neck oncology	Prospective	139 CI	GPT-4	2-item score	GPT cancer detection AC	71.2%
	Laser images	Cross-sectional	3 experts	PO vs GPT		Expert cancer detection AC	88.5%

Abbreviations: AC, accuracy; AE, additional examination; AIPI, artificial intelligence performance instrument; CC, clinical cases; CI, clinical images; cTNM, clinical tumor node metastasis; GPT, Generative Pre-Trained Transformer; OCAT, Ottawa clinical assessment tool; OTO, otolaryngologists; PD/DD, primary diagnosis/differential diagnoses; PO, practitioner in otolaryngology; TT, treatment.

Dallari et al, ChatGPT-3.5 accuracy was not influenced by the level of difficulty of the clinical case.[51] Moreover, authors observed that ChatGPT had difficulties in the management of some challenging diseases, particularly laryngopharyngeal reflux, which corroborated the data of another patient information study.[35] Finally, only 1 study explored the performance of ChatGPT-4 in the analysis of clinical images.[54] Thus, Sievert et al found that ChatGPT-4 and experts detected cancer at the microscopy in 71.2% and 88.5% of cases, respectively.[54]

Implications for Practice

Many aspects of otolaryngology practice have the potential to be revolutionized by generative AI, including the preparation of the consultation, the identification of potential emergency, the analysis of mucosa lesions, and the improvement of research ideas and protocols. To date, most studies focused on ChatGPT since the publicized launch of GPT-4.0 in March 2023. The present review reports that most applications of ChatGPT are limited in generating disease or treatment information or in the improvement of the management of clinical cases. The lack of comparison of ChatGPT performance with other LLMs is the main limitation of the current research. In the current otolaryngology literature, ChatGPT was compared with other LLMs, including Google Bard and others, in only 4 studies, which reported different results.[16,21,29,36] The landscape changes, and the continuous evolution of LLM's versions and performance make difficult the comparison between studies. The few studies in otolaryngology and beyond that exist have consistently demonstrated the superiority of ChatGPT-4.0 to other LLMs.[29] In the same vein, the stability of LLMs through regenerated responses is important for ensuring quality, consistency, and security in the provided information. However, the stability of ChatGPT has been explored in only a few studies, which reported conflicting results.[2,47-50,52] Dallari et al reported that the intraclass correlation coefficients for the stability of the correctness and consistency of ChatGPT-4 responses were 0.763 and 0.837, respectively.[47] In the study by Teixeira-Marques et al, the regenerated responses of ChatGPT-3.5 regarding the management of 20 reality-inspired clinical cases were similar in 15 cases.[48] The stability of regenerated responses of ChatGPT-4 was moderate-to-high in other studies dedicated to the management of head and neck,[50] and rhinology[52] cases, and in providing scientific references.[2] The rapid evolution of ChatGPT versions and the related changes in the landscape underscore the need for ongoing evaluation.

Moreover, compared to other applications of LLMs in the general medical literature, we found no applications of ChatGPT in medical education beyond its performance on professional exams. Given the implications of LLMs on learning and education, this finding highlights the urgent need for more scholarship in an aspect of otolaryngology that is bound to change with the rise of LLMs.

Another significant potential GPT application in otolaryngology that is not studied is its possible impact on clinical workflows, from electronic medical record search and summarization to ambient AI scribes and adverse event detection. This could have a significant impact on decreasing cognitive and work burden on otolaryngologists and is an important focus for investigation. Finally, ChatGPT's ability to analyze clinical images through the combination of LLM technology with image encoders has not yet been investigated in *Otolaryngology–Head and Neck Surgery*. This point is a key research topic for the future because most clinical diagnoses performed in otolaryngological practice are based on clinical or imaging findings. The analysis of images by multimodal LLM is an ongoing topic of research in other specialties. For instance, a recent study in ophthalmology tested the capabilities of ChatGPT-4 in answering ophthalmic cases from a medical education platform, including multiple-choice questions with or without ophthalmic imaging. The authors found that ChatGPT-4.0 performed better questions that did not rely on the interpretation of imaging modalities.[55] The findings of the present review may help otolaryngologists to propose new ideas of applications in further research.

The ethical aspect of the use of generative AI, such as ChatGPT, in research or clinical practice is another important point that requires further recommendations in *Otolaryngology–Head and Neck Surgery*. As it stands, entering patient information into the commercial version of ChatGPT infringes Health Insurance Portability and Accountability Act rule and should be avoided. As of January 24, 2024, the European Commission has proposed a legal framework through a positional statement paper to address the risks generated by specific uses of AI, especially in health care, which highlights the need to strengthen efforts to regulate the use of AI, while emphasizing the importance of safe, transparent, and human-centered use of AI technologies.[56] In the United States, President Biden's Executive Order on Safe, Secure, and Trustworthy AI has delineated new standards for safety and security, as well as the promotion of equity, civil right, consumer and worker protection, and innovation.[57]

Author Contributions

Jérôme R. Lechien, design, acquisition of data, data analysis, and interpretation, drafting, final approval, and accountability for the work, final approval of the version to be published, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved; **Anais Rameau**, design, acquisition of data, data analysis, and interpretation, revising the manuscript for important intellectual content, final approval of the version to be published, final approval, and accountability for the work, final approval of the version to be published, agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Disclosures

Competing interests: None.

Funding source: None.

References

1. Vaira LA, Lechien JR, Abbate V, et al. Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: a multicenter collaborative analysis. *Otolaryngol Head Neck Surg*. Published online August 18, 2023. doi:10.1002/ohn.489
2. Lechien JR, Briganti G, Vaira LA. Accuracy of ChatGPT-3.5 and -4 in providing scientific references in otolaryngology-head and neck surgery. *Eur Arch Otrhinolaryngol*. 2024;281:2159-2165. doi:10.1007/s00405-023-08441-8
3. Hoch CC, Wollenberg B, Lüers JC, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. *Eur Arch Otrhinolaryngol*. 2023;280(9):4271-4278. doi:10.1007/s00405-023-08051-4
4. Chiesa-Estomba CM, Lechien JR, Vaira LA, et al. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *Eur Arch Otrhinolaryngol*. 2023;281:2081-2086. doi:10.1007/s00405-023-08104-8
5. Lechien JR, Gorton A, Robertson J, Vaira LA. Is ChatGPT-4 accurate in proofread a manuscript in otolaryngology-head and neck surgery? *Otolaryngol Head Neck Surg*. Published online September 17, 2023. doi:10.1002/ohn.526
6. Novet J. Microsoft's new Windows 11 update adds a button promoting its Bing A.I. chatbot. February 28, 2023. Accessed January 20, 2024. https://www.cnbc.com/2023/02/28/microsoft-windows-11-update-adds-bing-ai-chatbot-button.html
7. Chiesa-Estomba CM, Speth MM, Mayo-Yanez M, Liu DT, Maniaci A, Borsetto D. Is the evolving role of artificial intelligence and chatbots in the field of otolaryngology embracing the future? *Eur Arch Otrhinolaryngol*. 2023;281:2179-2180. doi:10.1007/s00405-023-08382-2
8. Park I, Joshi AS, Javan R. Potential role of ChatGPT in clinical otolaryngology explained by ChatGPT. *Am J Otolaryngol*. 2023;44(4):103873. doi:10.1016/j.amjoto.2023.103873
9. Lechien JR. Generative AI and otolaryngology–head & neck surgery. *Otolaryngol Clin North Am*. 2024. Jun 4:S0030-6665(24)00064-1. doi:10.1016/j.otc.2024.04.006.
10. Alfertshofer M, Hoch CC, Funk PF, et al. Sailing the seven seas: a multinational comparison of ChatGPT's performance on medical licensing examinations. *Ann Biomed Eng*. Published online August 8, 2023. doi:10.1007/s10439-023-03338-3
11. Mahajan AP, Shabet CL, Smith J, Rudy SF, Kupfer RA, Bohm LA. Assessment of artificial intelligence performance on the Otolaryngology Residency In-Service Exam. *OTO Open*. 2023;7(4):e98. doi:10.1002/oto2.98

12. Long C, Lowe K, Zhang J, et al. A novel evaluation model for assessing ChatGPT on otolaryngology-head and neck surgery certification examinations: performance study. *JMIR Med Educ*. 2024;10:e49970. doi:10.2196/49970
13. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-analyses: the PRISMA statement. *PLoS Med*. 2009; 6(6):e1000097.
14. Zalzal HG, Cheng J, Shah RK. Evaluating the current ability of ChatGPT to assist in professional otolaryngology education. *OTO Open*. 2023;7(4):e94. doi:10.1002/oto2.94
15. Nielsen JPS, von Buchwald C, Grønhøj C. Validity of the large language model ChatGPT (GPT4) as a patient information source in otolaryngology by a variety of doctors in a tertiary otorhinolaryngology department. *Acta Otolaryngol*. 2023; 143(9):779-782. doi:10.1080/00016489.2023.2254809
16. Ayoub NF, Lee YJ, Grimm D, Divi V. Head-to-head comparison of ChatGPT versus Google search for medical knowledge acquisition. *Otolaryngol Head Neck Surg*. Published online August 2, 2023. doi:10.1002/ohn.465
17. Ayoub NF, Lee YJ, Grimm D, Balakrishnan K. Comparison between ChatGPT and Google search as sources of post-operative patient instructions. *JAMA Otolaryngol Head Neck Surg*. 2023;149(6):556-558. doi:10.1001/jamaoto.2023.0704
18. Shen SA, Perez-Heydrich CA, Xie DX, Nellis JC. ChatGPT vs. web search for patient questions: what does ChatGPT do better? *Eur Arch Otrhinolaryngol*. 2024;281(6):3219-3225.
19. Langlie J, Kamrava B, Pasick LJ, Mei C, Hoffer ME. Artificial intelligence and ChatGPT: an otolaryngology patient's ally or foe? *Am J Otolaryngol*. 2024;45(3):104220. doi:10.1016/j.amjoto.2024.104220
20. Moise A, Centomo-Bozzo A, Orishchak O, Alnoury MK, Daniel SJ. Can ChatGPT guide parents on tympanostomy tube insertion? *Children*. 2023;10(10):1634. doi:10.3390/children10101634
21. Bellinger JR, De La Chapa JS, Kwak MW, Ramos GA, Morrison D, Kesser BW. BPPV Information on Google versus AI (ChatGPT). *Otolaryngol Head Neck Surg*. Published online August 25, 2023. doi:10.1002/ohn.506
22. Capelleras M, Soto-Galindo GA, Cruellas M, Apaydin F. ChatGPT and rhinoplasty recovery: an exploration of ai's role in post-operative guidance. *Facial Plast Surg*. Published online January 10, 2024. doi:10.1055/a-2219-4901
23. Soto-Galindo GA, Capelleras M, Cruellas M, Apaydin F. Effectiveness of chatgpt in identifying and accurately guiding patients in rhinoplasty complications. *Facial Plast Surg*. Published online December 27, 2023. doi:10.1055/a-2218-6984
24. Durairaj KK, Baker O, Bertossi D, et al. Artificial intelligence versus expert plastic surgeon: comparative study shows ChatGPT "Wins" rhinoplasty consultations: should we be worried? *Facial Plast Surg Aesthetic Med*. Published online November 20, 2023. doi:10.1089/fpsam.2023.0224
25. Lim B, Seth I, Kah S, et al. Using generative artificial intelligence tools in cosmetic surgery: a study on rhinoplasty, facelifts, and blepharoplasty procedures. *J Clin Med*. 2023;12(20):6524. doi:10.3390/jcm12206524
26. Campbell DJ, Estephan LE, Sina EM, et al. Evaluating ChatGPT responses on thyroid nodules for patient education. *Thyroid*. 2024;34:371-377. doi:10.1089/thy.2023.0491
27. Mira FA, Favier V, Dos Santos Sobreira Nunes H, et al. Chat GPT for the management of obstructive sleep apnea: do we have a polar star? *Eur Arch Otrhinolaryngol*. 2023;281:2087-2093. doi:10.1007/s00405-023-08270-9
28. Campbell DJ, Estephan LE, Mastrolonardo EV, Amin DR, Huntley CT, Boon MS. Evaluating ChatGPT responses on obstructive sleep apnea for patient education. *J Clin Sleep Med*. 2023;19(12):1989-1995. doi:10.5664/jcsm.10728
29. Cheong RCT, Unadkat S, Mcneillis V, et al. Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: ChatGPT versus Google Bard. *Eur Arch Otrhinolaryngol*. 2024;281(2):985-993. doi:10.1007/s00405-023-08319-9
30. Davis RJ, Ayo-Ajibola O, Lin ME, et al. Evaluation of oropharyngeal cancer information from revolutionary artificial intelligence chatbot. *Laryngoscope*. 2023;134:2252-2257. doi:10.1002/lary.31191
31. Ostrowska M, Kacala P, Onolememen D, et al. "To trust or not to trust: evaluating the reliability and safety of AI responses to laryngeal cancer queries." Can our patients trust artificial intelligence? *Eur Arch Otorhinolaryngol*. Published online April 23, 2024.
32. Lee JC, Hamill CS, Shnayder Y, Buczek E, Kakarala K, Bur AM. Exploring the role of artificial intelligence chatbots in preoperative counseling for head and neck cancer surgery. *Laryngoscope*. Published online December 21, 2023. doi:10.1002/lary.31243
33. Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? *Front Oncol*. 2023;13:1256459. doi:10.3389/fonc.2023.1256459
34. Chiesa-Estomba CM, Urazan JD, Andueza M, et al. Comparative analysis of patient's perception between medical expert vs Chat-GPT advice for laryngeal, oropharyngeal, and salivary gland tumors. Clinical Presentation, San Sebastian University, San Sebastian, Spain, December 2023; 2024. In press.
35. Lechien JR, Carroll TL, Huston M, Naunheim MR. ChatGPT-4 accuracy for patient education in laryngopharyngeal reflux. *Laryngoscope*. 2024;281(5):2547-2552.
36. Cheong RCT, Pang KP, Unadkat S, et al. Performance of artificial intelligence chatbots in sleep medicine certification board exams: ChatGPT versus Google Bard. *Eur Arch Otrhinolaryngol*. 2023;281:2137-2143. doi:10.1007/s00405-023-08381-3
37. Lechien JR, Vaezi MF, Chan WW, et al. The Dubai Definition and Diagnostic Criteria of Laryngopharyngeal Reflux: The IFOS Consensus. *Laryngoscope*. 2023;134:1614-1624. doi:10.1002/lary.31134
38. Workman AD, Rathi VK, Lerner DK, Palmer JN, Adappa ND, Cohen NA. Utility of a LangChain and OpenAI GPT-powered chatbot based on the international consensus statement on allergy and rhinology: rhinosinusitis. *Int Forum Allergy*

Rhinol. Published oniline December 18, 2023. doi:10.1002/alr.23310

39. Koçak Z, Altay S. Balkan Medical Journal policy on the use of chatbots in scientific publications. *Balkan Med J*. 2023;40:149-150. doi:10.4274/balkanmedj.galenos.2023.17042023
40. Meyer JG, Urbanowicz RJ, Martin PCN, et al. ChatGPT and large language models in academia: opportunities and challenges. *BioData Min*. 2023;16:20. doi:10.1186/s13040-023-00339-9
41. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care*. 2023;27(1):75. doi:10.1186/s13054-023-04380-2
42. Frosolini A, Franz L, Benedetti S, et al. Assessing the accuracy of ChatGPT references in head and neck and ENT disciplines. *Eur Arch Otrhinolaryngol*. 2023;280(11):5129-5133. doi:10.1007/s00405-023-08205-4
43. Karimov Z, Allahverdiyev I, Agayarov OY, Demir D, Almuradova E. ChatGPT vs UpToDate: comparative study of usefulness and reliability of Chatbot in common clinical presentations of otorhinolaryngology-head and neck surgery. *Eur Arch Otrhinolaryngol*. 2024;281:2145-2151. doi:10.1007/s00405-023-08423-w
44. Qu RW, Qureshi U, Petersen G, Lee SC. Diagnostic and management applications of ChatGPT in structured otolaryngology clinical scenarios. *OTO Open*. 2023;7(3):e67. doi:10.1002/oto2.67
45. Saibene AM, Allevi F, Calvo-Henrique C, et al. Reliability of large language models in managing odontogenic sinusitis clinical scenarios: a preliminary multidisciplinary evaluation. *Eur Arch Otorhinolaryngol*. 2024;281:1835-1841. doi:10.1007/s00405-023-08372-4
46. Chee J, Kwa ED, Goh X. "Vertigo, likely peripheral": the dizzying rise of ChatGPT. *Eur Arch Otrhinolaryngol*. 2023;280(10):4687-4689. doi:10.1007/s00405-023-08135-1
47. Dallari V, Sacchetto A, Saetti R, Calabrese L, Vittadello F, Gazzini L. Is artificial intelligence ready to replace specialist doctors entirely? ENT specialists vs ChatGPT: 1-0, ball at the center. *Eur Arch Otorhinolaryngol*. 2024;281:995-1023. doi:10.1007/s00405-023-08321-1
48. Teixeira-Marques F, Medeiros N, Nazaré F, et al. Exploring the role of ChatGPT in clinical decision-making in otorhinolaryngology: a ChatGPT designed study. *Eur Arch Otrhinolaryngol*. 2024;281:2023-2030.
49. Marchi F, Bellini E, Iandelli A, Sampieri C, Peretti G. Exploring the landscape of AI-assisted decision-making in head and neck cancer treatment: a comparative analysis of NCCN guidelines and ChatGPT responses. *Eur Arch Otorhinolaryngol*. 2024;281:1-14. doi:10.1007/s00405-024-08525-z
50. Lechien JR, Chiesa-Estomba CM, Baudouin R, Hans S. Accuracy of ChatGPT in head and neck oncological board decisions: preliminary findings. *Eur Arch Otrhinolaryngol*. 2023;281:2105-2114. doi:10.1007/s00405-023-08326-w
51. Lechien JR, Georgescu BM, Hans S, Chiesa-Estomba CM. ChatGPT performance in laryngology and head and neck surgery: a clinical case-series. *Eur Arch Otrhinolaryngol*. 2023;281:319-333. doi:10.1007/s00405-023-08282-5
52. Radulesco T, Saibene AM, Michel J, Vaira LA, Lechien JR. ChatGPT-4 performance in rhinology: a clinical case series. *Int Forum Allergy Rhinol*. Published online January 24, 2024. doi:10.1002/alr.23323
53. 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 Otrhinolaryngol*. 2023;281:2063-2079. doi:10.1007/s00405-023-08219-y
54. Sievert M, Aubreville M, Mueller SK, et al. Diagnosis of malignancy in oropharyngeal confocal laser endomicroscopy using GPT 4.0 with vision. *Eur Arch Otrhinolaryngol*. 2024;281:2115-2122.
55. Mihalache A, Huang RS, Popovic MM, et al. Accuracy of an artificial intelligence chatbot's interpretation of clinical ophthalmic images. *JAMA Ophthalmol*. 2024;142:321.
56. European Commission. *A Strategic Vision to Foster the Development and Use of Lawful, Safe and Trustworthy Artificial Intelligence Systems in the European Commission*. European Commission; 2024.
57. The White House. The Biden-Harris Administration. Consulted February 19, 2024. 2024. Accessed January 20, 2024. https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence