European Archives of Oto-Rhino-Laryngology (2024) 281:6167–6172
https://doi.org/10.1007/s00405-024-08859-8

SHORT COMMUNICATION

Accuracy of ChatGPT responses on tracheotomy for patient education

Amina Khaldi[1] · Shahram Machayekhi[1] · Michele Salvagno[2] · Antonino Maniaci[3] · Luigi A. Vaira[4] · Luigi La Via[5] · Fabio S. Taccone[2] · Jerome R. Lechien[6,7,8]

Received: 17 July 2024 / Accepted: 18 July 2024 / Published online: 2 October 2024

Abstract

Objective To investigate the accuracy of information provided by ChatGPT-4o to patients about tracheotomy.

Methods Twenty common questions of patients about tracheotomy were presented to ChatGPT-4o twice (7-day intervals). The accuracy, clarity, relevance, completeness, referencing, and usefulness of responses were assessed by a board-certified otolaryngologist and a board-certified intensive care unit practitioner with the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool. The interrater reliability and the stability of the ChatGPT-4o responses were evaluated with intraclass correlation coefficient (ICC) and Pearson correlation analysis.

Results The total scores of QAMAI were 22.85 ± 4.75 for the intensive care practitioner and 21.45 ± 3.95 for the otolaryngologist, which consists of moderate-to-high accuracy. The otolaryngologist and the ICU practitioner reported high ICC (0.807; 95%CI: 0.655–0.911). The highest QAMAI scores have been found for clarity and completeness of explanations. The QAMAI scores for the accuracy of the information and the referencing were the lowest. The information related to the post-laryngectomy tracheostomy remains incomplete or erroneous. ChatGPT-4o did not provide references for their responses. The stability analysis reported high stability in regenerated questions.

Conclusion The accuracy of ChatGPT-4o is moderate-to-high in providing information related to the tracheotomy. However, patients using ChatGPT-4o need to be cautious about the information related to tracheotomy care, steps, and the differences between temporary and permanent tracheotomies.

Keywords ChatGPT · LLM · Artificial intelligence · Tracheotomy · Intensive care · Otolaryngology · Head neck · Surgery · Information · Accuracy

Pr Lechien and Pr Taccone have similarly contributed to the paper and can be joined as co-senior authors.

✉ Jerome R. Lechien
Jerome.Lechien@umons.ac.be

[1] Intensive Care Unit, EpiCURA Hospital, Hornu, Belgium

[2] Intensive Care Unit, Erasme Hospital, Brussels, Belgium

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

[4] Maxillofacial Surgery Operative Unit, Department of Medicine, Surgery and Pharmacy, University of Sassari, Viale San Pietro 43/B, Sassari 07100, Italy

[5] Department of Anesthesia and Intensive Care, University Hospital Policlinico "G.Rodolico-San Marco", Catania, Italy

[6] Department of Surgery, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium

[7] Department of Otolaryngology, Elsan Polyclinic of Poitiers, Poitiers, France

[8] Department of Otolaryngology-Head Neck Surgery, School of Medicine, UFR Simone Veil, Foch Hospital, Université Versailles Saint-Quentin-en-Yvelines (Paris Saclay University), Paris, France

Table 1 Quality Analysis of Medical Artificial Intelligence score of otolaryngologist

Questions
1. Can you cite the two main types of tracheotomies that are performed in ICU ?
2. What are the 5 most common indications of tracheotomy in ICU ?
3. What are the steps of the percutaneous tracheotomy ?
4. What are the steps of the surgical tracheotomy ? (1,2)
5. Is the tracheotomy useful for a patient with prolonged intubation; and what are the criteria ?
6. Why the tracheotomy is important for an intubated patient in ICU ?
7. What are the general criteria to transfer a patient from the ICU to a hospitalization department ?
8. What are the criteria to transfer a patient with tracheotomy from the ICU to
a hospitalization department ? (3)
9. In case of pandemic, what is the role/place/benefit of tracheotomy ?
10. Can the tracheotomy accelerate the rehabilitation of patients?
11. Are there several cannula for tracheotomy and what are their differences? (4)
12. Can a tracheotomized patient speak?
13. What is the role of the balloon associated with the cannula ?
14. For patients requiring long-term tracheotomy, is there an alternative device ? (5)
15. Can you cite 3 clinical conditions associated with the need to have permanent tracheotomy ?
16. What are the common care associated with the tracheotomy ?
17. What are the advantages of the various cannula types ?
18. What are the features of the tracheostomy in laryngectomized patients ?
19. What are the differences between tracheotomy (non laryngectomized)
and tracheostomy (laryngectomized patient) ? (6)
Errors
1. Mention the need of a suture to close the skin orifice of the tracheotomy at the end of the procedure.
2. Mention the need for a sterile bandage at the end of tracheotomy to reduce the risk of infection
3. ChatGPT does not mention the importance of training/skills of nursing in the management of
tracheotomy in the hospitalization unit.
4. ChatGPT does not mention/develop the fenestrated cannula, and the larytube for laryngectomized patients.
5. ChatGPT does not mention the Montgomery tube, and proposes a prosthesis with electronic voice.
6. ChatGPT does not mention the definitive suture between trachea and skin for laryngectomized
patient and states that both types of tracheotomy terms are similar.

Table 1 footnotes: The questions associated with a ChatGPT error are identified with a number under brackets. Abbreviations: ICU = intensive care unit.

Introduction

The development of artificial intelligence-powered language models, such as Chatbot Generative Pre-trained Transformer (ChatGPT) is emerging in medicine.[1] ChatGPT is one of the most used chatbots by patients and practitioners, leading to an increase in the number of studies dedicated to the evaluation of the accuracy of information provided by ChatGPT [1, 2]. The last version (ChatGPT-4o) is available for patients who wish to collect information about medicine and surgery. To date, many authors have investigated the accuracy of ChatGPT in patient counseling, guidance, and education in general otolaryngology, [3, 4] pediatric otolaryngology, [5] otology, [6] rhinology, [7] endocrine otolaryngology, [8] sleep, [9, 10] laryngology, [11] and head and neck oncology [12]. Tracheotomy is a common procedure in otolaryngology-head and neck surgery, as well as in the intensive care unit (ICU). In this way, patients might be tempted to interrogate Chatbot to obtain information about the indication, steps, and complications of the procedure.

This study aimed to investigate the accuracy of ChatGTP-4o in providing information related to tracheotomy, which was not previously performed.

Methods

Questions and setting

Twenty questions related to tracheotomy indication, steps, and complications were collected by four practitioners, including two board-certified otolaryngologists-head and neck surgeons (J.R.L.,and A.M.), and two board-certified intensive care unit practitioners (A.K., and S.S.). The list of questions was based on common questions of patients' families and patients themselves in case of indication of tracheotomy (Table 1). The questions were related to the types and features of tracheotomy procedures ($N=3$); indications and timing ($N=4$); steps of procedures ($N=2$); types and characteristics of cannula ($N=5$); care, rehabilitation, and follow-up ($N=5$); usefulness of tracheotomy in a pandemic situation ($N=1$); and post-tracheotomy rehabilitation ($N=1$). The questions were entered twice (7-day intervals) in the ChatGPT-4o (Open AI, San Francisco, USA) interface in June 2024, which is accessible through the application programming interface (API; https://chat.openai.com). The responses were collected in a document available for judges.

Accuracy evaluation

The primary outcomes of the study were the evaluation of accuracy, clarity, relevance, completeness, referencing, and usefulness of responses were independently assessed by a board-certified otolaryngologist (J.R.L.) and a board-certified intensive care unit practitioner (A.K.) using the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool (Fig. 1) [13]. QAMA is a validated and standardized tool dedicated to the evaluation of Chatbot accuracy. QAMAI was developed by an international collaborative group of experts who come from the Young Otolaryngologists of the International Federation of Otorhinolaryngological Society. In practice, the QAMAI tool has been developed based on the Modified DISCERN (mDISCERN) instrument [14]. The mDISCERN is a well-validated and widely used tool for assessing the quality of health information conveyed by websites, social networks/media, YouTube and other multimedia platforms with each parameter rating with a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The score was summed into an overall score (QAMAI score), which determines the quality of the information [13]. The secondary outcomes consisted of the collection of the number and the types of errors of ChatGPT-4o.

Statistical methods

Statistical analyses were performed using the Statistical Package for the Social Sciences for Windows (SPSS version 30,0; IBM Corp, Armonk, NY, USA). The accuracy, clarity, relevance, completeness, referencing, usefulness, and total QAMAI scores related to the ChatGPT-4o responses were reported with mean and standard deviations. The interrater reliability of judges was investigated with the intraclass correlation coefficient (ICC). The stability of ChatGTP-4o responses was evaluated with the Pearson correlation coefficient. The consistency was considered as low, moderate, and strong for $k < 0.40$, 0.40–0.60, and $k > 0.60$, respectively. A level of significance of $p < 0.05$ was used.

Results

The QAMAI scores of ChatGPT-4o responses reported by the otolaryngologist and the intensive care unit practitioner are reported in Table 2. The total scores of QAMAI were 22.85 ± 4.75 and 21.45 ± 3.95 for the ICU physician and the otolaryngologist, respectively ($p = 0.265$), which consists of moderate-to-high accuracy. The highest performances of ChatGPT-4o were found for clarity of explanations, and completeness (Table 2). The lowest performances were related to the accuracy of the information, and the lack of references. The errors of ChatGPT-4o to tracheotomy questions were reported in Table 1. These errors were related to the surgical steps of tracheotomy, the lack of consideration of the training of the nursing team (tracheotomy care), the types of cannulas, and the distinction between temporary and permanent tracheotomies. Precisely, the information related to the post-laryngectomy tracheostomy remains incomplete and erroneous. The stability of ChatGPT-4o responses is described in Table 3. The stability of responses was high for all items. The otolaryngologist and the ICU practitioner reported high ICC (0.807; 95%CI: 0.655–0.911). The responses of ChatGPT-4o are available in Appendix 1.

	1 Strongly Disagree	2 Disagree	3 Neutral	4 Agree	5 Strongly Agree
Accuracy: The information provided is accurate and up-to-date					
Clarity: The answer is clear and comprehensible in terms of language and scientific terminology.					
Relevance: the information provided is relevant and directly answer to the question posed.					
Completeness: the response adequately covers all aspects of the question and provides sufficient information including areas of uncertainty.					
Provision of sources and references: the response provides reliable sources and references to support the health information presented					
Usefulness: the response provides to meet the user's health information needs.					

Fig. 1 The Quality Analysis of Medical Artificial Intelligence tool. *Notes*: Each item is assessed with a 5-point Likert scale (Strongly disagree, disagree, neutral, agree, strongly agree)

Table 2 Quality Analysis of Medical Artificial Intelligence score of otolaryngologist and intensive care practitioner

	ICU Practitioner					
QAMAI items (5-Likert scale)	Types	Indications	Steps	Cannula	Care/rehab	Total
Accuracy	3.67 ± 1.53	3.75 ± 1.26	4.50 ± 0.71	3.80 ± 1.64	3.40 ± 1.34	3.65 ± 1.31
Clarity	5.00 ± 0.01	5.00 ± 0.01	5.00 ± 0.01	5.00 ± 0.01	5.00 ± 0.01	5.0 ± 0.01
Relevance	3.67 ± 1.53	4.50 ± 0.58	5.00 ± 0.01	3.60 ± 1.52	3.60 ± 1.14	3.85 ± 1.23
Completeness	3.33 ± 1.16	4.00 ± 0.82	4.50 ± 0.71	3.60 ± 1.52	3.80 ± 1.30	3.70 ± 1.17
Referencing	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01
Usefulness	3.33 ± 1.16	4.00 ± 0.82	5.00 ± 0.01	3.60 ± 1.52	3.40 ± 1.34	3.65 ± 1.22
QAMAI total score (/30)	22.00 ± 5.29	24.25 ± 2.87	27.00 ± 1.41	22.60 ± 6.15	22.20 ± 5.02	22.85 ± 4.75
	OTO-HNS					
QAMAI items (5-Likert scale)	Types	Indications	Steps	Cannula	Care/rehab	Total
Accuracy	2.67 ± 2.08	4.50 ± 1.00	3.00 ± 0.01	3.20 ± 1.10	3.40 ± 0.89	3.45 ± 1.19
Clarity	4.00 ± 1.73	4.50 ± 0.58	4.50 ± 0.71	4.20 ± 0.84	4.40 ± 0.55	4.35 ± 0.81
Relevance	2.67 ± 1.16	4.75 ± 0.50	4.00 ± 0.01	3.20 ± 0.84	3.80 ± 0.45	3.70 ± 0.92
Completeness	2.67 ± 1.16	4.00 ± 0.82	3.50 ± 0.71	3.20 ± 0.84	3.60 ± 0.55	3.45 ± 0.83
Referencing	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01	3.00 ± 0.01
Usefulness	2.67 ± 1.16	4.00 ± 0.82	3.50 ± 0.71	3.00 ± 1.00	4.00 ± 0.71	3.50 ± 0.95
QAMAI total score	17.67 ± 6.66	24.75 ± 3.40	21.50 ± 0.71	19.80 ± 3.27	22.20 ± 2.28	21.45 ± 3.95

Table 2 footnotes: Abbreviations: ICU = intensive care unit; OTO-HNS = otolaryngologist head and neck surgeon; QAMAI = Quality Analysis of Medical Artificial Intelligence Score.

Table 3 Stability of ChatGPT-4o responses

QAMAI items (5-Likert scale)	Pearson	*p*-value
Accuracy	0.826	0.001
Clarity	1.000	0.001
Relevance	0.804	0.001
Completeness	0.830	0.001
Referencing	1.000	0.001
Usefulness	0.829	0.001
QAMAI total score (/30)	0.840	0.001

Table 3 footnotes: Abbreviations: ICU = intensive care unit; QAMAI = Quality Analysis of Medical Artificial Intelligence Score.

Discussion

The number of tracheostomies performed annually in resource-rich countries is estimated at 250,000 with 10% carried out in children [15] In the United States and the United Kingdom, over 100,000 and 15,000 people receive tracheotomies every year, [15] making the tracheotomy one of the most common surgical procedures worldwide. In this way, patients or patients' families commonly search for information related to tracheotomy online with ChatGPT as an emerging resource for such information [1].

In the present study, ChatGPT-4o reports moderate-to-high levels of accuracy, clarity, relevance, and completeness for the information related to the tracheotomy, which corroborates the findings of the literature for common otolaryngological conditions and procedures [16–19]. Depending on the method used to evaluate the accuracy, 57–89% of information provided by ChaGPT-3.5 or 4.0 were judged as accurate for education, medical advice, or understandability [16–19]. Interestingly, the accuracy of ChatGPT's information can be influenced by the complexity of the topic [1]. Thus, the responses dedicated to general disorders or surgery, such as tonsillectomy, functional endoscopic sinus surgery, or tympanoplasty reported the highest accuracy, [1, 19' while the information provided by ChatGPT-4 for rare and complicated disorders, such as bilateral vocal cord paralysis, reported the lowest accuracy [1]. The high prevalence and popularity of tracheotomy can support the performance of ChatGPT-4o in providing accurate information.

However, some mistakes have been identified in the answer provided by ChatGPT-4o, and they were related to details about the tracheotomy procedure, the post-operative care, and the differences between temporary and permanent tracheotomies. These mistakes were related to technical or specific points, which can be considered as less important for patient understanding. The mistakes for technical or specific points were similarly reported by Campbell et al. who found 2–28% of mistakes by ChatGPT-3.5 in providing information related to thyroid and endocrine conditions [8]. These authors also found that ChatGPT-3.5 did not provide accurate references for supporting the responses [8]. In the present study, ChatGPT-4o did not provide scientific references in all questions, which limits authors in the referencing's analysis. The performance of ChatGPT-4o in providing references was investigated in some studies, [1, 2, 20] which reported mixed conclusions. Thus, Vaira et al. reported that ChatGPT-4 provided 50% of false references to support some answers, [20] whereas Langlie et al. did not find major errors or hallucinations [19].

The stability of the responses provided by an artificial intelligence-powered language model is an important

parameter for ensuring quality, consistency, and security in the provided information.[1] In the present study, ChatGPT-4o responses were stable when the questions were regenerated in the API. Our results corroborate the findings of Kuscu et al. who found 94.1% of similar responses provided by ChatGPT-4 in regenerated questions related to head and neck surgical procedures [21].

The primary strengths of this study are its originality and the consideration of both otolaryngologist and intensive care unit practitioners for the analysis of ChatGPT-4o responses. To date, there is no study investigating the accuracy of information provided by ChatGPT for tracheotomy, which remains a common procedure performed in medicine and surgery. The evaluation of the stability of the responses is an additional strength.

The lack of inclusion of patients to evaluate the clarity of ChatGPT-4o responses is the primary limitation of the study. Indeed, the clarity of information was evaluated by healthcare professionals, while the opinion of patients could bring additional information for the overall judgment of the ChatGPT-4o performance.

Conclusions

The accuracy of ChatGPT-4o is moderate-to-high in providing information related to the tracheotomy. However, patients using ChatGPT-4o need to be cautious about the information related to tracheotomy care, steps, and the differences between temporary and permanent tracheotomies.

Acknowledgements None.

Author contributions J.R. Lechien: design, acquisition of data, data analysis & 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. A. Khaldi: design, acquisition of data, data analysis & 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. M. Salvagno, S. Mashayekhi, A. Maniaci, L. Vaira, L. La Via: design, data analysis & interpretation, revising the manuscript for important intellectual content. F.S. Taccone: 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.

Funding None.

Declarations

Competing interests The author Jerome R. Lechien is also guest editor of the special issue on 'ChatGPT and Artificial Intelligence in Otolaryngology—Head and Neck Surgery'. He was not involved with the peer review process of this article.

Sponsorships None.

Disclosure of potential conflicts of interest Authors have no conflict of interest.

Informed consent Not required.

Ethic committee Not required.

References

1. Lechien JR, Rameau A (2024) Applications of ChatGPT in Otolaryngology-Head Neck surgery: a state of the Art Review. Otolaryngol Head Neck Surg. https://doi.org/10.1002/ohn.807
2. Lechien JR (2024) Generative AI and Otolaryngology-Head & Neck Surgery. *Otolaryngol Clin North Am.* : S0030-6665(24)00064–1. https://doi.org/10.1016/j.otc.2024.04.006
3. Abou-Abdallah M, Dar T, Mahmudzade Y, Michaels J, Talwar R, Tornari C (2024) The quality and readability of patient information provided by ChatGPT: can AI reliably explain common ENT operations? Eur Arch Otorhinolaryngol. https://doi.org/10.1007/s00405-024-08598-w
4. Tessler I, Wolfovitz A, Alon EE, Gecel NA, Livneh N, Zimlichman E, Klang E (2024) ChatGPT's adherence to otolaryngology clinical practice guidelines. Eur Arch Otorhinolaryngol 281(7):3829–3834. https://doi.org/10.1007/s00405-024-08634-9
5. Maniaci A, Lazzeroni M, Cozzi A, Fraccaroli F, Gaffuri M, Chiesa-Estomba C, Capaccio P (2024) Can chatbots enhance the management of pediatric sialadenitis in clinical practice? Eur Arch Otorhinolaryngol. https://doi.org/10.1007/s00405-024-08798-4
6. Bellinger JR, De La Chapa JS, Kwak MW, Ramos GA, Morrison D, Kesser BW (2023) BPPV Information on Google Versus AI (ChatGPT). Otolaryngol Head Neck Surg. https://doi.org/10.1002/ohn.506
7. Riestra-Ayora J, Vaduva C, Esteban-Sánchez J, Garrote-Garrote M, Fernández-Navarro C, Sánchez-Rodríguez C, Martin-Sanz E (2024) ChatGPT as an information tool in rhinology. Can we trust each other today? Eur Arch Otorhinolaryngol 281(6):3253–3259. https://doi.org/10.1007/s00405-024-08581-5
8. Campbell DJ, Estephan LE, Sina E, Mastrolonardo EV, Alapati R, Amin DR, Cottrill E (2023) Evaluating ChatGPT responses on thyroid nodules for patient education. Thyroid. https://doi.org/10.1089/thy.2023.0491
9. Mira FA, Favier V, Dos Santos Sobreira Nunes H, de Castro JV, Carsuzaa F, Meccariello G, Vicini C, De Vito A, Lechien JR, Estomba CC, Maniaci A, Iannella G, Rojas EP, Cornejo JB, Cammaroto G (2023) Chat GPT for the management of obstructive sleep apnea: do we have a polar star? Eur Arch Otorhinolaryngol. https://doi.org/10.1007/s00405-023-08270-9
10. Cheong RCT, Unadkat S, Mcneillis V, Williamson A, Joseph J, Randhawa P, Andrews P, Paleri V (2024) Artificial intelligence chatbots as sources of patient education material for obstructive sleep apnoea: ChatGPT versus Google Bard. Eur Arch Otorhinolaryngol 281(2):985–993. https://doi.org/10.1007/s00405-023-08319-9

11. Lechien JR, Carroll TL, Huston MN, Naunheim MR (2024) ChatGPT-4 accuracy for patient education in laryngopharyngeal reflux. Eur Arch Otorhinolaryngol 281(5):2547–2552. https://doi.org/10.1007/s00405-024-08560-w
12. Ostrowska M, Kacała P, Onolememen D, Vaughan-Lane K, Sisily Joseph A, Ostrowski A, Pietruszewska W, Banaszewski J, Wróbel MJ (2024) To trust or not to trust: evaluating the reliability and safety of AI responses to laryngeal cancer queries. Eur Arch Otorhinolaryngol. https://doi.org/10.1007/s00405-024-08643-8
13. 13.Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, Bergonzani M, Boscolo-Rizzo P, Califano G, Cammaroto G, Chiesa-Estomba CM, Committeri U, Crimi S, Curran NR, di Bello F, di Stadio A, Frosolini A, Gabriele G, Gengler IM, Lonardi F, Maglitto F, Mayo-Yáñez M, Petrocelli M, Pucci R, Saibene AM, Saponaro G, Tel A, Trabalzini F, Trecca EMC, Vellone V, Salzano G, De Riu G (2024) Validation of the Quality Analysis of Medical Artificial Intelligence (QAMAI) tool: a new tool to assess the quality of health information provided by AI platforms. Eur Arch Otorhinolaryngol. https://doi.org/10.1007/s00405-024-08710-0
14. Charnock D, Shepperd S, Needham G et al (1999) DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 53:105–111
15. Brenner MJ, Pandian V, Milliren CE, Graham DA, Zaga C, Morris LL, Bedwell JR, Das P, Zhu H, Lee Y, Allen J, Peltz A, Chin K, Schiff BA, Randall DM, Swords C, French D, Ward E, Sweeney JM, Warrillow SJ, Arora A, Narula A, McGrath BA, Cameron TS, Roberson DW, Lee YJ, Grimm D, Divi V (2020) Global Tracheostomy Collaborative: data-driven improvements in patient safety through multidisciplinary teamwork, standardisation, education, and patient partnership. *Br J Anaesth*. ; 125(1):e104-e118. https://doi.org/10.1016/j.bja.2020.04.054.
16. 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. https://doi.org/10.1080/00016489.2023.2254809.
17. 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*. 2023. https://doi.org/10.1002/ohn.465.
18. Shen SA, Perez-Heydrich CA, Xie DX, Nellis JC (2024) ChatGPT vs. web search for patient questions: what does ChatGPT do better? Eur Arch Otorhinolaryngol 281(6):3219–3225. https://doi.org/10.1007/s00405-024-08524-0
19. Langlie J, Kamrava B, Pasick LJ, Mei C, Hoffer ME (2024) Artificial intelligence and ChatGPT: an otolaryngology patient's ally or foe? Am J Otolaryngol 45(3):104220. https://doi.org/10.1016/j.amjoto.2024.104220
20. Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, Bergonzani M, Bolzoni A, Committeri U, Crimi S, Gabriele G, Lonardi F, Maglitto F, Petrocelli M, Pucci R, Saponaro G, Tel A, Vellone V, Chiesa-Estomba CM, Boscolo-Rizzo P, Salzano G, De Riu G (2024) Accuracy of ChatGPT-Generated information on Head and Neck and Oromaxillofacial surgery: a Multicenter Collaborative Analysis. Otolaryngol Head Neck Surg 170(6):1492–1503. https://doi.org/10.1002/ohn.489
21. Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S (2023) Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? Front Oncol 13:1256459. https://doi.org/10.3389/fonc.2023.1256459