Are ChatGPT’s Answers to Questions About Salivary Gland Diseases Accurate and Reliable?

Original Investigation

Zülküf Küçüktağ, Esma Altan, Gökçe Saygı Uysal, Aykut Özdoğan

University of Health Sciences Türkiye, Ankara Etlik City Hospital, Department of Otorhinolaryngology-Head and Neck Surgery, Ankara, Türkiye

Abstract

Objective: Chat Generative Pre-Trained Transformer (ChatGPT) is an artificial intelligence model that can generate human-like text dialogs to inputs. There are no ChatGPT studies in the literature only on salivary gland diseases. This study evaluates the accuracy and reliability of ChatGPT’s answers to questions in salivary gland diseases, focusing on its potential use in training otolaryngology professionals.

Methods: Sixty-one questions, categorized as “basic knowledge,” and “salivary gland tumors” were posed twice using ChatGPT-4. Answers were categorized as 1 (completely correct and comprehensive), 2 (partially correct), 3 (misleading information containing correct and incorrect statements), or 4 (completely incorrect). The accuracy of the answers was evaluated by two ear, nose and throat specialists. Inconsistencies in the rating were resolved by a third reviewer. Reproducibility was assessed by the agreement between the first and second answers.

Results: Fifteen (24.6%) of the questions were about basic knowledge, while 46 (75.4%) were about salivary gland tumors. ChatGPT gave “completely correct and comprehensive” answers to 54 (88.5%) questions and “partially correct” answers to seven (11.5%) questions. “Misleading information containing correct and incorrect statements” and “completely incorrect” answers were not received. The reproducibility rate of first and second answers was 97%.

Conclusion: ChatGPT provided highly accurate and reproducible answers to questions about salivary gland diseases. ChatGPT is thought to be an important source of information for otolaryngology professionals. Although the results of our study show that ChatGPT is highly successful, more studies are needed in this field.

Keywords: Salivary gland, salivary gland tumor, sialadenitis, generative artificial intelligence, medical education, professional practice

ORCID IDs of the authors:

Z.K. 0000-0001-6846-3771
E.A. 0000-0002-3080-3571
G.S.U. 0000-0001-9704-3522
A.Ö. 0009-0001-6460-3383

Cite this article as: *Küçüktağ Z, Altan E, Saygı Uysal G, Özdoğan A. Are ChatGPT’s answers to questions about salivary gland diseases accurate and reliable? Turk Arch Otorhinolaryngol. [Epub Ahead of Print]*

Corresponding Author:
Zülküf Küçüktağ, MD;
zkucuk79@yahoo.com

Received Date: 13.02.2025
Accepted Date: 30.06.2025
Epub: 22.09.2025

DOI: 10.4274/tao.2025.2025-1-17

Introduction

In recent years, developments in artificial intelligence (AI) models have started to have an impact on many areas of life. The integration of AI into medical fields affects human health. One of these AI applications is the Chat Generative Pre-Trained Transformer (ChatGPT), which is an interactive chat engine, and a large language model trained with internet text data.

ChatGPT is an AI model capable of generating human-like conversational dialogue by generating answers to questions from a large knowledge database (1-4). ChatGPT, a large language model developed by Open AI and trained on internet-based data, is one of the important developments in AI (5). ChatGPT, which has a wide range of information sources, can generate human-like responses to text and sentence inputs (1-3). It is also capable of providing information on various topics, answering questions, and chatting. It can do so in both medical and non-medical fields (6). Since applications and information in the medical field require high responsibility and transparency, it is of great importance to develop an AI system with accurate and reliable medical knowledge (7). The need for both healthcare professionals, and medical students, as well as patients, to obtain information from ChatGPT makes the reliability of this application even more important. Its performance improves with continuous and repetitive inputs, in other words, with user interaction (6).

The use of ChatGPT, especially in the medical field, has brought some controversies. While some researchers consider the medical information provided by ChatGPT as valuable, others have distanced themselves from this issue due to misuse during medical writing, security issues, accuracy of information, and legal concerns (8,9).

ChatGPT serves as an additional source of information that otolaryngology professionals can use for their academic training and exam preparation. In the field of otolaryngology, AI studies have been reported on clinical staging methods, analyzing cochlear implant performance, detection of parathyroid gland, prediction of prognosis in otolaryngology and head and neck surgery patients, determination of accuracy and reliability of information about head and neck cancers (4,10-13). Our study is unique in that it only included questions about salivary gland diseases. The aim of this study is to determine the accuracy and reliability of ChatGPT's answers to questions about salivary gland diseases, and thereby to determine whether the AI application can be used as a resource in the training of otolaryngology professionals on this subject.

Methods

Study Design

The GPT-4 version of ChatGPT (OpenAI, San Francisco, CA) was used for the study. ChatGPT was asked a total of 61 questions about salivary glands. The questions were developed based on standard otolaryngology textbooks, clinical guidelines, and the authors' clinical experience. They were not formatted as examination questions but were designed to reflect clinically relevant scenarios that can be encountered in practice. All questions were prepared by an ear, nose and throat (ENT) specialist with over ten years of clinical experience. To prevent bias, the evaluation of ChatGPT's answers was conducted by different ENT specialists who were not involved in the question preparation process.

The questions were systematically divided into two different groups: basic knowledge and salivary gland tumors. To evaluate the consistency and reproducibility of ChatGPT's answers and to reduce memory bias, each question was asked twice on the same day, one after the other, from the same computer using the "new input" function. Thus, each answer for the same question was reproduced twice and scored independently. All questions asked to ChatGPT were asked in English and the questions and answers received were archived (Supplementary File). Since our study was not a study involving humans and animals, ethics committee approval and patient consent were not required.

Grading System

Two ENT specialists who were actively working, experienced in their field (more than 10 years of experience), and who did not communicate with each other about the questions independently reviewed and graded the ChatGPT answers (first and second) for accuracy and reproducibility. The accuracy of the answers was determined by the scoring method of Kuşcu et al. (4):

1. Comprehensive/correct: Completely correct and comprehensive data

2. Incomplete/partially correct: Partially correct data

3. Mixed: Misleading information containing correct and incorrect statements

4. Completely inaccurate/irrelevant: Completely incorrect data

Reproducibility was assessed and scored independently by two ENT specialists according to the consistency of the two answers from ChatGPT to each question. If the two answers were similar, only the first answer given by ChatGPT was recorded and scored. If the answers were different, both answers were scored and recorded by the ENT specialists. Both ENT specialists had more than 10 years of experience and were actively involved in both clinical and academic studies. No residents or junior doctors participated in the evaluation process.

When the scores of the first and second answers given by ChatGPT were different, the answers were considered not reproducible, i.e., incongruent. All discrepancies in the accuracy and reproducibility of answers between the two reviewers were reviewed and resolved by a third experienced ENT specialist (with more than 10 years of experience) who was blinded to the initial reviews.

Statistical Analysis

Statistical analysis of the data was performed with IBM SPSS Statistics for Windows, version 27 (IBM Corp., Armonk, NY, USA). Descriptive statistics were calculated as number, percentage, mean, standard deviation, median and min-max. Inter-measurement consistencies were evaluated by intraclass correlation coefficient (ICC). In the evaluation of ICC coefficients below 0.4 was considered poor, between 0.4-0.59 moderate, between 0.60-0.74 good and above 0.75 excellent relationship. P-values less than 0.05 were considered statistically significant (Mann-Whitney U test). The Wilcoxon sign test was performed to study whether there was a statistically significant difference between the answers obtained from the questions asked to ChatGPT two times.

Results

A total of 61 questions were asked to ChatGPT about salivary gland diseases. Fifteen (24.6%) of the questions were about basic knowledge and non-tumor diseases of the salivary glands, while 46 (75.4%) were about salivary gland tumors. The distribution of ChatGPT answers to both question groups is shown in Table 1. ChatGPT gave "completely correct and comprehensive" answers to 54 (88.5%) questions and "partially correct" answers to seven (11.5%) questions. None of the questions were scored as "misleading information containing correct and incorrect statements" or "completely incorrect." These results are shown graphically in Figure 1.

The agreement of ChatGPT's answers to questions first and second, in other words reproducibility, was 96.7% (59 out of 61 questions). This rate was 100% for basic knowledge questions and 95.6% for salivary gland tumors questions. These results are shown graphically in Figure 2. There was no statistically significant difference between the first and second answers given by ChatGPT (p=0.157) (Table 2). In general, the first answers of ChatGPT were more accurate than the second answers. A total of 91.8% of the first answers and 88.5% of the second answers were evaluated as completely accurate and comprehensive.

ICC was used to examine the consistency of the decisions made by the reviewers who evaluated the answers given by ChatGPT. Accordingly, there was no statistically significant difference between the scores given by the reviewers to the questions in both groups (p=0.779). For the basic knowledge questions, the ICC rate was 1.00, indicating high consistency. When the answers related to salivary gland tumors were evaluated by two reviewers, ICC was found to be 0.899 for ChatGPT's first answers and 0.959 for second answers, indicating a high degree of consistency between the two reviewers (Table 3).

Table 1. Distribution of answers received from ChatGPT according to question groups

	Number of questions (%)
Basic knowledge (n=15)	
Completely correct and comprehensive	15 (100)
Partially correct	-
Misleading information containing correct and incorrect statements	-
Completely incorrect	-
Salivary gland tumors (n=46)	
Completely correct and comprehensive	39 (84.8)
Partially correct	7 (15.2)
Misleading information containing correct and incorrect statements	-
Completely incorrect	-

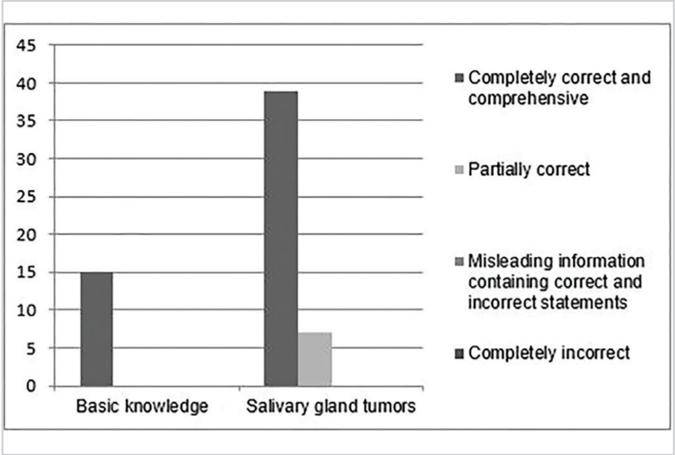

Figure 1. Graphical representation of the answers provided by ChatGPT according to the question categories

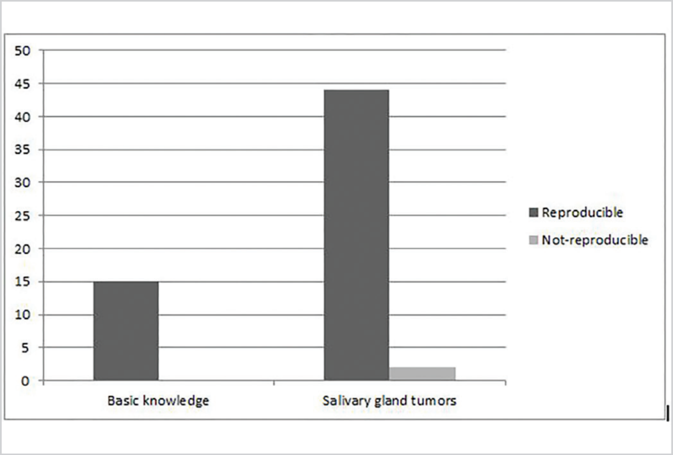

Figure 2. Reproducibility of answers according to the question categories

Table 2. Concordance analysis of 1st and 2nd answers from ChatGPT

Answers	Reproducibility	Reviewers compliance	
	n (%)	1st answers, n (%)	2nd answers, n (%)
Basic knowledge	15 (100)	15 (100)	15 (100)
p-value	**1.00**	**1.00**	**1.00**
Salivary gland tumors	44 (95.6)	45 (97.8)	45 (97.8)
p-value	**0.157**	**0.779**	**0.779**
Total	59 (96.7)	60 (98.3)	60 (98.3)
p-value	**0.157**	**0.779**	**0.779**

*p<0.05

Table 3. Accuracy rate of answers from ChatGPT according to reviewers

			Completely correct and comprehensive n (%)	Partially correct n (%)	Misleading information containing correct and incorrect statements n (%)	Completely incorrect n (%)	p-value	ICC
Basic knowledge	1st answers	1	15 (100)				1.00	1.00
		2	15 (100)					
	1st answers	1	15 (100)				1.00	1.00
		2	15 (100)					
Salivary gland tumors	1st answers	1	41 (89.1)	5 (10.9)			0.538	0.899
		2	39 (84.8)	7 (15.2)				
		3	41 (89.1)	5 (10.9)				
	2nd answers	1	39 (84.8)	7 (15.2)			0.779	0.959
		2	38 (82.6)	8 (17.4)				
		3	39 (84.8)	7 (15.2)				
Total	1st answers	1	56 (91.8)	5 (8.2)			0.545	0.902
		2	54 (88.5)	7 (11.5)				
		3	56 (91.8)	5 (8.2)				
	2nd answers	1	54 (88.5)	7 (11.5)			0.784	0.961
		2	53 (86.9)	8 (13.1)				
		3	54 (88.5)	7 (11.5)				

ICC: Intraclass correlation coefficient, p: Mann-Whitney U test p-value

(1: First reviewer / 2: Second reviewer / 3: Third reviewer)

Discussion

In this study, the accuracy and reproducibility of ChatGPT's answers to questions about salivary gland diseases were found to be acceptably high. However, the study was aimed solely at ENT healthcare professionals. There were no questions intended to inform patients or their relatives.

The use of AI models in various fields, especially in healthcare, is increasing. Since ChatGPT was introduced to the market in November 2022, it has become an important source of information, especially for healthcare professionals, to access medical information related to their field. It is important for both clinicians and patients to note that not every response received from ChatGPT should be taken as medical advice (4). Medical information obtained from AI models should not be considered as direct information, but as a reference that directs us to primary information (4). Park et al. (14) stated that ChatGPT helped clinicians in their decision-making processes; however, since the model has the potential to give erroneous information and is mostly based on previously taught data, its limitations should be well known and used carefully in view of patient safety. In a study supporting this statement, ChatGPT gave completely correct answers to eight out of 20 questions about allergic rhinitis for patient education, five partially correct answers and the remaining questions with varying degrees of misinformation. In this study, it was pointed out that it may be risky to rely completely on chatbots such as ChatGPT for medical advice. It was reported that patients should always seek the opinions of health professionals and online resources should only be used as a complementary information tool. ChatGPT and similar chatbots may be useful for patient education, but they can never replace healthcare professionals (15).

Since it is a new application, there are limited number of studies on the use of ChatGPT in otolaryngology. A literature review reveals two studies involving ChatGPT on salivary glands. However, in one of these studies, the effectiveness of ChatGPT in sialendoscopy decision-making was evaluated, not its ability to provide information about salivary gland diseases. For this, ChatGPT answers were compared with the decisions of 10 expert sialendoscopists. A statistically significant agreement was found between ChatGPT and sialendoscopists and it was concluded that ChatGPT was a promising model for clinical decision making, especially for patients suitable for sialendoscopy treatment (16). Another study was conducted by Hoch et al. (6) with questions about 15 sub-branches of otolaryngology, including 138 questions about salivary glands. Our study was a ChatGPT study that included only questions about salivary glands. The questions consisted of two separate question groups: questions about basic knowledge about salivary gland diseases and questions about clinical approach including diagnosis, treatment and management of salivary gland tumors. The purpose of preparing the questions in two different categories was to compare ChatGPT's answers to the more easily accessible basic knowledge and clinical-based questions that are considered to be relatively more difficult and comprehensive. In fact, it is thought that ChatGPT's success in basic medical sciences questions is higher than in clinical-based questions that require making a diagnosis by interpreting the symptoms. This is because information on basic medical sciences can be accessed directly in the literature. In our study, in support of this information, the accuracy rate of ChatGPT on basic knowledge about salivary gland diseases was higher than that of questions about tumors. The findings of the study by Seifen et al. (17) also support this statement. Seifen et al. (17) compared the answers of ChatGPT and a certified specialist in sleep disorders on the interpretation of polysomnography results and treatment recommendations for sleep apnea. There was 97% agreement between ChatGPT and the sleep specialist in the diagnosis of simple cases and 100% agreement in treatment recommendations. In patients with positive airway pressure intolerance, there was 70% agreement between ChatGPT and the sleep specialist in diagnosis and 44% agreement in treatment recommendations. Consistent with our findings, ChatGPT performs better on basic questions, whereas its success rate decreases for more complex topics such as treatment management.

One of the remarkable results of our study is that "misleading information containing correct and incorrect statements" and "completely incorrect" answers were not received to any of the questions. Kuşcu et al. (4) investigated the accuracy and reliability of ChatGPT answers to 154 questions about head and neck cancers. ChatGPT answered "completely correct and comprehensive" to 86.4% (133/154) of the questions. The rates for "partially correct" and "misleading information containing correct and incorrect statements" were 11% and 2.6%, respectively, and no "completely incorrect" answers were received. To evaluate the performance and reproducibility of ChatGPT, Tessler et al. (18) repeatedly asked ChatGPT 24 clinical otolaryngology questions based on the American Academy of Otolaryngology guidelines. While 59.7% (43/72) of the answers were completely correct, only 2.8% (2/72) were incorrect.

The ChatGPT study with the largest question archive in the field of otolaryngology is the study conducted by Hoch et al. (6) with 2,576 questions (479 multiple-choice and 2,097 single-choice questions) on 15 different sub-branches of otolaryngology. When ChatGPT's answers to these questions were evaluated, 57% of the questions were answered correctly. ChatGPT had the highest number of correct answers to allergy questions (72%), and the lowest number of correct answers to questions related to legal otolaryngology (29%). In this study, there were 138 single-answer multiple-choice questions related to salivary glands, and ChatGPT answered 60.9% of these 138 questions correctly and 39.1% incorrectly. Compared to these studies, the accuracy rate of ChatGPT answers was found to be quite high in our study.

Open-ended questions aim to simulate real-life clinical scenarios that clinicians often encounter and assess clinicians' judgement and ability to draw conclusions (19). Some of our questions included real case scenarios. "What should be the surgical approach when intraoperative facial nerve invasion is encountered in a patient with a malignant parotid tumor who had no preoperative signs of facial paralysis?" or "What does it mean if a pathology report for malignant salivary glands includes the term lymphovascular and/or perineural invasion?" are examples to such questions. ChatGPT was observed to be very successful in these questions. The questions in our study and in the study of Kuşcu et al. (4) were open-ended questions. When the results of these two studies were compared with those of Hoch et al. (6) it was observed that ChatGPT was more successful with open-ended questions rather than single-answer multiple-choice questions. The results of the study by Zalzal et al. (20) also support this statement. In the study by Zalzal et al. (20), ChatGPT was first asked 30 open-ended questions and then 30 single-answer multiple-choice questions about otolaryngology and the answers were checked by two experienced ENT specialists. In the open-ended questions, the ChatGPT model initially gave 56.7% completely correct and 86.7% partially correct answers. When the questions were repeated, the model increased to 73.3% completely correct and 96.7% partially correct. However, ChatGPT performed significantly worse on single-answer multiple-choice questions, with only 43.3% correct answers. When answering open-ended questions, it may be sufficient to give general information about the subject. However, single-answer multiple-choice questions are not based on interpretation and may require knowledge of the finest detail about the subject.

Kuşcu et al. (4) included questions in their study on head and neck cancers that were designed to inform both healthcare professionals and patients/patient relatives. In contrast, the questions in our study were exclusively aimed at healthcare professionals and did not include questions intended to inform patients or their relatives. It is thought that more accurate and adequate results can be obtained by conducting more studies on salivary gland diseases, improving the areas of use of ChatGPT, adding up-to-date information, and improving the database. In addition, studies investigating the accuracy and reliability of the information that patients and their relatives will obtain from ChatGPT on this subject should also be conducted.

In our study, the fact that three different experienced ENT specialists evaluated independently of each other enabled us to avoid examiner-induced errors and biases. In addition, the fact that there was no statistically significant difference between the scores given by the examiners and that the ICC rate was high for the answers received increased the reliability of the results. It is thought that asking each question to ChatGPT separately, rather than asking two questions in one sentence, will increase reliability and accuracy.

In the study of Kuşcu et al. (4) the reproducibility rate between the answers of ChatGPT was found 94.1%. In another study, the agreement between the answers of the model was found 70.8% and it was stated that there was a reasonable consistency between the answers (18). Lechien and Rameau (21) reported that ChatGPT was a helpful model for editing scientific manuscripts, preparing study protocols, preparing student and assistant exams, and that the consistency of the answers given to repetitive questions in these subjects was high.

Despite its positive aspects, there are also studies showing that ChatGPT has significant shortcomings and needs to be improved over time. The best examples are the studies by Karimov et al. (22) and Hoch et al. (6). In the study by Karimov et al. (22) in which ChatGPT was compared with the UpToDate search engine, it was shown that UpToDate provided more accurate and reliable answers to the findings of 25 different clinical scenarios in the field of otorhinolaryngology than ChatGPT, and that UpToDate, unlike ChatGPT, supported the information it provided with tables, figures and algorithms. Hoch et al. (6) stated that ChatGPT can be a supplementary resource in otolaryngology examinations, but it needs to be further improved due to its error tendency and lack of knowledge in some areas of otolaryngology. Another study in which concerns were expressed about the use of ChatGPT in otolaryngology education was conducted by Long et al. (19). Twenty-one open-ended questions were taken from the sample exam of the Royal College of Physicians and Surgeons of Canada and asked to ChatGPT-4. ChatGPT-4 was successful in this exam with a passing grade. However, the success rate of the answers increased when clues were given. In addition, the fact that some of the answers given were incorrect and contradictory was considered a worrying situation. As a result, it was suggested that additional adjustments should be made to obtain more reliable and accurate answers for clinical practice, as they may provide erroneous information that may threaten patient safety. It is deemed important to integrate ChatGPT into a broader learning strategy. Information from AI models should be supported by textbooks, lectures, and training with subject experts. This combination provides a better learning experience and alleviates potential credibility and ethical concerns regarding the use of AI models alone for educational purposes (6). The fact that ChatGPT compiles information from other sources may cause the information accessed about different sub-branches of medicine to have different limits. The best example of this is expressed in the article by Hoch et al. (6) who stated that the category of legal issues with the lowest accuracy rate referred to German medical laws and the database used in this field could be more limited, which posed a challenge for ChatGPT. On the other hand, the higher correct response rates in some sub-branches of otolaryngology were attributed to the wider data sources and comprehensive pools of accessible information. In addition, topics with a high rate of correct answers, such as allergy, may be topics that ChatGPT users frequently search for medical advice. This is interpreted as regular user interaction improving the performance and the accuracy of the model (6).

In conclusion, although ChatGPT has some shortcomings and despite the concerns, it will continue to be an important source of information in the field of otolaryngology. The fact that it has high accuracy and reproducibility rates in some subjects, as in our study, shows that AI models are promising.

Study Limitations

The use of open-ended questions in this study allowed for more detailed responses; however, it also led to a limitation in the total number of questions that could be included. The study could be further developed by organizing questions under specific subtopics and expanding the question pool. Additionally, the evaluation was limited to text-based responses only, without considering ChatGPT's ability to interpret visual data in medical decision-making. This represents a gap in assessing the model's potential for clinical applications. Lastly, this study focused solely on the educational use of ChatGPT and did not include questions directed at patients or their relatives. Future research that encompasses a wider range of topics and includes visual elements is expected to provide more comprehensive contributions to literature.

Conclusion

The accuracy, reliability and reproducibility of ChatGPT-4 responses related to salivary gland diseases were found to be high. It is considered a reliable resource for healthcare professionals, otolaryngology residents and students. Further studies are needed to improve its role in clinical decision-making.

Ethics

Ethics Approval: Since our study was not a study involving humans and animals, ethics committee approval were not required.

Informed Consent: Patient consent were not required.

Footnotes

Authorship Contributions

Concept: Z.K., E.A., Design: Z.K., E.A., G.S.U., A.Ö., Data Collection and/or Processing: Z.K., E.A., G.S.U., A.Ö., Analysis and/or Interpretation: Z.K., E.A., G.S.U., A.Ö., Literature Search: Z.K., E.A., G.S.U., A.Ö., Writing: Z.K., E.A.

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Disclosure: The authors declare that this study has received no financial support.

Main Points

- ChatGPT answered 88.5% of the questions "completely correct and comprehensive." None of the answers were scored as "misleading information containing correct and incorrect statements" or "completely incorrect."
- The rate of ChatGPT's answers to repeated questions, i.e., reproducibility, was 96.7%. In other words, the answers obtained by asking the same question again were found to be compatible with each other.
- Considering the above data, ChatGPT can be a reliable additional resource for otolaryngology professionals on salivary gland diseases.

References

1. Munoz-Zuluaga C, Zhao Z, Wang F, Greenblatt MB, Yang HS. Assessing the accuracy and clinical utility of ChatGPT in laboratory medicine. Clin Chem. 2023; 69: 939-40. [Crossref]
2. The Lancet Digital Health. ChatGPT: friend or foe? Lancet Digit Health. 2023; 5: e102. [Crossref]
3. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. Nature. 2023; 614: 224-6. [Crossref]
4. 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. [Crossref]
5. Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, et al. Accuracy of ChatGPT-generated information on head and neck and oromaxillofacial surgery: a multicenter collaborative analysis. Otolaryngol Head Neck Surg. 2024; 170: 1492-1503. [Crossref]
6. Hoch CC, Wollenberg B, Lüers JC, Knoedler S, Knoedler L, Frank K, 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 Otorhinolaryngol. 2023; 280: 4271-8. [Crossref]
7. Masters K. Artificial intelligence in medical education. Med Teach. 2019; 41: 976-80. [Crossref]

8. Else H. Abstracts written by ChatGPT fool scientists. Nature. 2023; 613: 423. [Crossref]

9. Haupt CE, Marks M. AI-generated medical advice-GPT and Beyond. JAMA. 2023; 329: 1349-50. [Crossref]

10. Knoedler L, Baecher H, Kauke-Navarro M, Prantl L, Machens HG, Scheuermann P, et al. Towards a reliable and rapid automated grading system in facial palsy patients: facial palsy surgery meets computer science. J Clin Med. 2022; 11: 4998. [Crossref]

11. Crowson MG, Dixon P, Mahmood R, Lee JW, Shipp D, Le T, et al. Predicting postoperative cochlear implant performance using supervised machine learning. Otol Neurotol. 2020; 41: e1013-23. [Crossref]

12. Wang B, Zheng J, Yu JF, Lin SY, Yan SY, Zhang LY, et al. Development of artificial intelligence for parathyroid recognition during endoscopic thyroid surgery. Laryngoscope. 2022; 132: 2516-23. [Crossref]

13. Qu RW, Qureshi U, Petersen G, Lee SC. Diagnostic and management applications of ChatGPT in structured otolaryngology clinical scenarios. OTO Open. 2023; 7: e67. [Crossref]

14. Park I, Joshi AS, Javan R. Potential role of ChatGPT in clinical otolaryngology explained by ChatGPT. Am J Otolaryngol. 2023; 44: 103873. [Crossref]

15. Høj S, Thomsen SF, Meteran H, Sigsgaard T, Meteran H. Artificial intelligence and allergic rhinitis: does ChatGPT increase or impair the knowledge? J Public Health (Oxf). 2024; 46: 123-6. [Crossref]

16. Chiesa-Estomba CM, Lechien JR, Vaira LA, Brunet A, Cammaroto G, Mayo-Yanez M, et al. Correction: exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. Eur Arch Otorhinolaryngol. 2024; 281: 2777. [Crossref]

17. Seifen C, Huppertz T, Gouveris H, Bahr-Hamm K, Pordzik J, Eckrich J, et al. Chasing sleep physicians: ChatGPT-4o on the interpretation of polysomnographic results. Eur Arch Otorhinolaryngol. 2025; 282: 1631-9. [Crossref]

18. Tessler I, Wolfovitz A, Alon EE, Gecel NA, Livneh N, Zimlichman E, et al. ChatGPT's adherence to otolaryngology clinical practice guidelines. Eur Arch Otorhinolaryngol. 2024; 281: 3829-34. [Crossref]

19. Long C, Lowe K, Zhang J, Santos AD, Alanazi A, O'Brien D, et al. A novel evaluation model for assessing ChatGPT on otolaryngology-head and neck surgery certification examinations: performance study. JMRI Med. Educ. 2024; 10: e49970. [Crossref]

20. Zalzal HG, Cheng J, Shah RK. Evaluating the current ability of ChatGPT to assist in professional otolaryngology education. OTO Open. 2023; 7: e94. [Crossref]

21. Lechien JR, Rameau A. Applications of ChatGPT in otolaryngology-head neck surgery: a state of the art review. Otolaryngol Head Neck Surg. 2024; 171: 667-77. [Crossref]

22. 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 Otorhinolaryngol. 2024; 281: 2145-51. [Crossref]

Supplementary File:
https://d2v96fxpocvxx.cloudfront.net/34c1fd7d-947b-4954-9ae2-39560c57d146/content-images/a2c51588-86fc-40a9-b404-8cf4022c509f.pdf