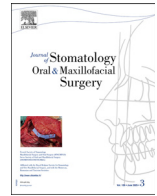




Available online at
ScienceDirect
 www.sciencedirect.com

Elsevier Masson France
EM|consulte
 www.em-consulte.com



Original Article

ChatGPT and trainee performances in the management of maxillofacial patients



Mélissa Peters^{a,*}, Maxime Le Clercq^a, Antoine Yanni^a, Xavier Vanden Eynden^a,
 Lalmand Martin^a, Noémie Vanden Haute^a, Szonja Tancredi^a, Céline De Passe^a,
 Edward Boutremans^a, Jerome Lechien^{b,c,d,e,f,g}, Didier Dequanter^{a,b}

^a Department of Stomatology, Oral & Maxillofacial Surgery, CHU Saint Pierre, Brussels, Belgium

^b Faculty of Medicine, Department of Human Anatomy and Experimental Oncology UMONS, Mons, Belgium

^c Phonetics and Phonology Laboratory (UMR 7018 CNRS, Université Sorbonne Nouvelle/Paris 3), Department of Otorhinolaryngology and Head and Neck Surgery, Foch Hospital, School of Medicine, UFR Simone Veil, Université Versailles Saint-Quentin-en-Yvelines (Paris Saclay University), Paris, France

^d Department of Otorhinolaryngology and Head and Neck Surgery, CHU Saint-Pierre, Brussels, Belgium

^e Research Committee of Young Otolaryngologists of the International Federation of Otorhinolaryngological Societies (IFOS), Paris, France

^f Young Confederation of the European Oto-Rhino-Laryngological Head and Neck Surgery Societies (Y-CEORLHNS), Dublin, Ireland

^g 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

ARTICLE INFO

Article History:

Received 23 June 2024

Accepted 22 September 2024

Available online 25 September 2024

Key-words:

ChatGPT

Comparison

Diagnosis

Maxillofacial surgery

ABSTRACT

Introduction: ChatGPT is an artificial intelligence based large language model with the ability to generate human-like response to text input, its performance has already been the subject of several studies in different fields. The aim of this study was to evaluate the performance of ChatGPT in the management of maxillofacial clinical cases.

Materials and methods: A total of 38 clinical cases consulting at the Stomatology-Maxillofacial Surgery Department were prospectively recruited and presented to ChatGPT, which was interrogated for diagnosis, differential diagnosis, management and treatment. The performance of trainees and ChatGPT was compared by three blinded board-certified maxillofacial surgeons using the AIPI score.

Results: The average total AIPI score assigned to the practitioners was 18.71 and 16.39 to ChatGPT, significantly lower ($p < 0.001$). According to the experts, ChatGPT was significantly less effective for diagnosis and treatment ($p < 0.001$). Following two of the three experts, ChatGPT was significantly less effective in considering patient data ($p = 0.001$) and suggesting additional examinations ($p < 0.0001$). The primary diagnosis proposed by ChatGPT was judged by the experts as not plausible and/or incomplete in 2.63 % to 18 % of the cases, the additional examinations were associated with inadequate examinations in 2.63 %, to 21.05 % of the cases and proposed an association of pertinent, but incomplete therapeutic findings in 18.42 % to 47.37 % of the cases, while the therapeutic findings were considered pertinent, necessary and inadequate in 18.42 % of cases.

Conclusions: ChatGPT appears less efficient in diagnosis, the selection of the most adequate additional examination and the proposition of pertinent and necessary therapeutic approaches.

© 2024 Elsevier Masson SAS. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

1. Introduction

More than 70 % of people in Europe, Asia and the USA consult the Internet to answer their medical questions [1]. With 1.8 billion visits per day in April 2023, ChatGPT is becoming a considerable source of information [1–3], and could be used by patients to answer medical

questions. Launched in 2020, ChatGPT is an artificial intelligence based (AI-based) large language model (LLM) trained on massive text datasets in multiple languages with the ability to generate human-like response to text input [1,4]. Developed by OpenAI (OpenAI,LLC., San Francisco,CA,USA), ChatGPT etymology is related to being a chat-bot (a program able to understand and generate response using a text-based interface) and is based on the generative pre-trained transformer (GPT) architecture[5,6]. Since then, many clinical studies have been conducted to evaluate the performance of ChatGPT in

* Corresponding author: CHU Saint Pierre, Rue Haute, 320, 1000 Brussels, Belgium.
 E-mail address: Melissa.peters@stpierre-bru.be (M. Peters).

different areas, such as scientific manuscript revisions, or in some clinical fields; for example, the use of ChatGPT was evaluated for finding references and editing language in medical scientific articles [7–9]. In the scientific community and academia, ChatGPT received mixed responses reflecting the history of controversy regarding the benefits vs. risk of advanced AI technologies [10]. Nowadays, most medical experts agree with the clinical potential usefulness of ChatGPT according to its large database. However, this new technology should be investigated for its capabilities and potential risks. Actually, some authors claimed that the reliability of the free version of ChatGPT (v.3.0) used in the diagnosis and the management of real clinical cases appears to be limited [11]. Moreover, the same authors reported that ChatGPT could not discern the superiority of some additional examinations over others and was not able to diagnose some atypical conditions in patients with complex medical or surgical histories. Other studies also support that ChatGPT is a good information tool for patients, but it's not accurate enough for professionals, it's still far from playing a reliable support role in clinical decision-making and is only a useful complement [2,3].

The objective of our study was to evaluate the performance of ChatGPT in the management of clinical cases consulting at the Stomatology-maxillofacial Department, from the development of diagnosis, differential diagnosis, suggestion of additional examinations to the treatment proposal, using the AIPI score.

2. Material and methods

2.1. Patients and setting

From September to November 2023, 38 patients were prospectively recruited from the Department of Stomatology-Maxillofacial Surgery of CHU Saint-Pierre (Brussels, Belgium). Patient medical records (history, symptoms, physical examination, differential diagnosis, additional examinations, and treatments) were recorded in a database by a single trainee in maxillofacial surgery. Incomplete

clinical cases were excluded. Then, these consultation findings were evaluated by three senior maxillofacial surgeons based on the current guidelines, and, therefore, considered as the standard of care (adequate management) for the assessment of the ChatGPT performance (AIPI score established following these guidelines). The guidelines consisted of the scientific position paper/recommendations of the European Society of Maxillofacial Surgery. A total AIPI score and sub-scores were established for each clinical case. The AIPI score was developed and validated by three experts from the International Federation of ENT Societies (YO-IFOS) [11]. The AIPI included 9 items assessing medical and surgical history; symptoms; physical examination; diagnosis; additional examinations; management plan, and treatments (Fig. 1). The final AIPI score ranges from 0 to 20 with a score of 20 indicating excellent clinical case management by the AI, while a score of 0 is associated with inadequate management. AIPI may be subdivided into the 4 following sub-scores associated with common items: patient feature score (/6), diagnosis score (/7), additional examination score (/4), and treatment score (/3). AIPI provides a comprehensive approach to clinical cases, intended for use in daily practice.

2.2. ChatGPT performance and accuracy

Afterwards, the data of the consultation were presented in november 2023 to ChatGPT (version 3.5), solely in text form, without images and without mentioning the human differential diagnoses, additional examinations, and treatments. ChatGPT was systematically interrogated in french with this 4 questions about for differential diagnoses (What are your differential diagnoses?), primary diagnoses (What is your primary diagnoses?), additional examinations (What are your additional examinations to find the diagnosis?), and potential therapeutic approach(es) (What are your treatment(s) for the primary diagnosis?). The ChatGPT findings were collected in a database. A total AIPI score and subscores were defined and compared with the expert's evaluation.

Outcomes of artificial intelligence performance instrument (AIPI)	Practitioner evaluation	Item score	Subscores
1. Consideration of medical and surgical history in the AI management	Fully (2) Partly (1) Not (0)	.../2	Patient feature score .../6
2. Consideration of symptoms of patients in the AI management	Fully (2) Partly (1) Not (0)	.../2	
3. Consideration of physical findings reported by practitioner(s)	Fully (2) Partly (1) Not (0)	.../2	
4. The differential diagnoses provided by AI are	Complete and plausible (3) Incomplete but plausible (2) Incomplete and not plausible for one or several (1) Absent (0)	.../3	Diagnosis score .../7
5. The primary diagnosis of AI was	Correct (3) Plausible (2) Not plausible (1) Absent (0)	.../3	
6. The management plan of AI included potential physical/additional examinations for determining the diagnosis	Yes (1) No (0)	.../1	
7. The additional examinations proposed by AI are/include	All pertinent and necessary examinations (3) All pertinent but partially necessary examinations (2) As association of pertinent, necessary and inadequate examinations (1) An association of inadequate examinations (0)	.../3	Additional examination score .../4
8. AI identified the most relevant additional examination to perform first	Yes (1) No (0)	.../1	
9. The treatments proposed by AI are/include	No, AI provided a list without stratification (0) All pertinent and necessary therapeutic findings (3) All pertinent but incomplete therapeutic findings (2) As association of pertinent, necessary, and inadequate therapeutic findings (1) No adequate therapeutic findings (0)	.../3	Treatment score .../3
		Total AIPI	... /20

Fig. 1. Artificial intelligence performance instrument. AIPI score ranges from 0 (inadequate management) to 20 (adequate management).

The local ethics committee approved the study protocol (CHUSP, n°BE0762023230708). The patient consented to participate. *Statistical analyses*

Statistical analysis were performed through the Statistical Package for the Social Sciences for Windows (SPSS version 24.0; IBM Corp, Armonk, NY, USA). A level of significance of $p < 0.05$ was used. The interrater reliability (concordance analysis) was assessed by comparing the total AIPI scores and subscores of 3 experienced practitioners with Kendall's W (coefficient of concordance) and Friedman test. Kendall's W was used to assess the similarity between the expert's ratings of the AIPI scores. Friedman test was used to evaluate the similarity (absolute agreement) of the values of the scores given by the experts. Descriptive analyses and Wilcoxon rank analyze to compare AIPI scores obtained by the practitioners and ChatGPT were performed.

3. Results

A total of 38 patients (23 males and 15 females) completed the evaluation. The mean age was 35.5 years. The primary diagnosis and/or diagnostic hypothesis was made during the consultation in all patients. ChatGPT was interrogated for all patient cases.

In our series, among the diagnoses, infection was the most frequent clinical diagnosis ($n = 15$; 39.5 %) followed by trauma and mucosal pathologies ($N = 7$; 18.5 %) respectively (Table 1).

Medical history, symptoms, physical examination, additional examination, differential diagnosis, and treatment findings of patients are available in Appendix A1 and B1.

The average total AIPI score assigned to the practitioners by the three experts was 18.71. A less significant average total AIPI score determined by the experts of 16.39 was attributed to ChatGPT ($p < 0.001$). According to the experts, ChatGPT was significantly less effective for diagnosis and treatment($p < 0.001$). Following the evaluation of two of the three experts, ChatGPT was significantly less effective in considering patient data and suggesting additional examinations ($p = 0.01$) and $p < 0.0001$) respectively. Total and subscores of AIPI are reported in Table 2.

Interrater reliability did not report significant differences in the AIPI sub- and total scores among the three independent experts. According to subscores, the interrater reliability was 0.547 ($p < 0.008$)

According to the three experts (DD, ML and XVDE), the medical and surgical history were partially considered by ChatGPT in 5.26 %, 13.16 %, 39.47 % of the cases respectively. ChatGPT considered the symptoms partially in 2.63 %, 5.26 %, 7.89 % of the cases while the physical findings were partially considered by ChatGPT in 7.89 %, 10.53 %, 36.84 % according to the experts. The primary diagnosis was judged by the experts as not plausible in 7.89 %, 15.79 %, 18.42 % of the cases and incomplete in 2.63 %, 5.26 %, 7.89 % of the cases. Experts reported that additional examinations proposed by ChatGPT were associated with inadequate examinations in 2.63 %, to 21.05 % of the cases. The most relevant additional examination to perform first was chosen correctly in 7.89 %, 42.11 %, 60.53 % of the cases respectively. Regarding treatments, experts reported that ChatGPT proposed an association of pertinent, but incomplete therapeutic findings in 18.42 %, 39.47 % and 47.37 % of the cases, while the therapeutic findings

Table 1
Patient diagnosis.

38 clinical cases	
Age (average)	37.58 years old (3–86)
Sex (N, %)	23 M (60.52 %), 15 F (39.47 %)
Diagnosis (N, %)	
- Infection	15 (39.47 %)
- Facial trauma	7 (18.42 %)
- Salivary gland pathology	3 (7.89 %)
- Dental trauma	2 (5.26 %)
- Temporomandibular dysfunction	2 (5.26 %)
- Oncology	2 (5.26 %)
- Other (dental or mucosal pathology, post-op complications, etc.)	7 (18.42 %)

Table 2
ChatGPT performance and maxillofacial surgeons's evaluations.

AIPI score – practitioner			AIPI score – ChatGPT			P value
Senior 1	Senior 2	Senior 3	Senior 1	Senior 2	Senior 3	$P < 0.001$
17.79	18.76	19.58	16.47	16.21	16.5	
Average = 18.71			Average = 16.39			

were considered pertinent, necessary and inadequate in 18.42 % of cases.

4. Discussion

The rapid development of ChatGPT means that this technology is increasingly present in our consultations and discussions with our patients. However, according to the differences between Humans and machine assessment, the usability and reliability of such clinical instruments needs to be evaluated. In order to evaluate the clinical instrument, AIPI, dedicated to IA performance assessment, was developed [11,12]. The psychometric analyses support that AIPI is a valid and reliable clinical instrument for rating the performance of ChatGPT in the management of real clinical cases. The internal consistency, test–retest reliability, interrater reliability, and internal validity reported adequate values, which corroborate the findings of other clinical performance assessment tools [13–15].

In the present study, we observed that AIPI total scores were significantly higher in Humans compared to ChatGPT clinical case evaluation. The average total AIPI score assigned to the practitioners by three experts was 18.71. A less significant average total AIPI score determined by the three experts of 16.39 was attributed to ChatGPT ($p < 0.001$). Moreover, according to the experts, ChatGPT was significantly less effective for diagnosis and treatment($p < 0.001$). Following the evaluation of two of the three experts, ChatGPT was significantly less effective in considering patient data and suggesting additional examinations ($p = 0.01$) and $p < 0.0001$) respectively (Table 3). According to the experts, in our study, the medical and surgical history were partially considered by ChatGPT in 5.26 % to 39.47 % of the cases respectively. Moreover, ChatGPT considered the symptoms partially in 2.63 % to 7.89 % of the cases while the physical findings were partially considered by ChatGPT in 7.89 % to 36.84 % according to the experts. These results may suggest that the current version of

Table 3
Accuracy of ChatGPT judged by senior maxillofacial surgeons.

	Senior 1	Senior 2	Senior 3
Patient feature sub-score	No difference	Practitioner > ChatGPT $p = 0.01$	Practitioner > ChatGPT $p = 0.03$
Diagnosis sub-score	Practitioner > ChatGPT $p < 0.001$	Practitioner > ChatGPT $p = 0.01$	Practitioner > ChatGPT $p = 0.001$
Additional examination sub-score	No difference	Practitioner > ChatGPT $p < 0.0001$	Practitioner > ChatGPT $p < 0.0001$
Treatment sub-score	Practitioner > ChatGPT $p < 0.001$	Practitioner > ChatGPT $p < 0.001$	Practitioner > ChatGPT $p < 0.001$

ChatGPT functions more as an electronic encyclopedia providing a potential list of differential diagnoses and additional examinations, rather than a virtual practitioner considering the patient features. As an example, the suggestion to take paracetamol in a patient who had already consumed 15 g of paracetamol was ignored by ChatGPT (patient n°21, [appendix A1](#)). The responses generated by ChatGPT should therefore be treated with great caution and should not be relied upon by patients as a substitute for consultation with a health-care professional. As it has not been developed for use in the medical field, aberrant responses could have serious medical consequences for the patient.

Moreover, our findings correlated with the literature exploring the reliability of ChatGPT in medical practice and more specifically in the field of head and neck pathologies. In fact, the theoretical performance of ChatGPT in the field of maxillofacial pathologies has been evaluated first in radiology. Mago *et al.* demonstrated that ChatGPT-3.5 was efficient in describing the pathology, characteristics radiographic features, and describing anatomical landmarks. The authors concluded that, definitively, ChatGPT-3.5 could be used as an adjunct when an oral radiologist needs additional information on any pathology. However, the chatbot system cannot be the mainstay for reference [3]. Viara *et al.* confirmed, in their multicentric study, a good level of accuracy in the AI's answers. The authors emphasized the AI's ability to resolve complex clinical scenarios, but again the chatbot system failed to be considered as a reliable support for the decision-making process as demonstrated by our results [2]. Otherwise, the clinical utility of ChatGPT was supported by one recent study. Chiesa-Estomba *et al.* investigated the level of agreement between ChatGPT and 10 international sialendoscopists aiming the capabilities of ChatGPT to further improve the management of salivary gland disorders. The authors reported a significant agreement between ChatGPT and experts in the clinical decision-making process within the salivary gland clinic, which supports the theoretical performance of ChatGPT-3.5 [16]. In other medical disciplines and in particular in otorhinolaryngology, ChatGPT demonstrated potential as a valuable information resource. In their study, the authors emphasized the importance of AI-human collaboration, but they insisted that ChatGPT served as a complementary tool rather than a replacement for medical professionals [17,18].

In the same way, Qu *et al.* concluded that the performance of ChatGPT meets a high agreement rate for easy- to-moderate difficulty clinical cases with physicians for diagnosis and treatment [19]. Our study reached the same conclusions. The primary diagnosis was judged by the experts as not plausible in 7.89 % to 18.42 % of the cases and incomplete in 2.63 % to 7.89 % of the cases. Regarding treatments, experts reported that ChatGPT proposed an association of pertinent, but incomplete therapeutic findings in 18.42 % to 47.37 % of the cases, while the therapeutic findings were considered pertinent, necessary and inadequate in 18.42 % of cases. In head and neck oncological decisions [20], ChatGPT-4 provided adequate explanations in 19/20 cases (95 %). Chat -GPT proposed a higher number of additional examinations than the practitioners and the therapeutic propositions were accurate in 13 cases (65 %), leading to the same conclusions observed in our study.

In the literature, other studies explored the reliability of ChatGPT in other medical fields. In their study, Frosolini *et al.* explored the reliability of references generated by ChatGPT language models in the Head and Neck field. They demonstrated that ChatGPT displayed a tendency to provide erroneous/non-existent references [21]. These findings underlined that information/recommendations provided by ChatGPT need to be considered with precautions, keeping in mind that the human discernment of the practitioner is not yet acquired by chatbot systems. Hoch *et al.* evaluated the accuracy of ChatGPT's response to practice quiz questions designed for otolaryngology board certification [22]. The dataset included 2576 questions, of which 57 % ($n = 1475$) were answered correctly by ChatGPT. An in-

depth analysis of question style revealed that single-choice questions were associated with a significantly higher rate ($p < 0.001$) of correct responses ($n = 1313; 63\%$) compared to multiple-choice questions ($n = 162; 34\%$). These reported results are consistent with the clinical observations. Nielsen *et al.*, in their study, confirmed, that, despite its potential in providing relevant medical information, the chatbot's responses lacked depth and were found to potentially perpetuate biases [1]. The same may be applied to patients. Indeed, according to the mediatization of ChatGPT performance, it is conceivable that the number of patients who will use the chatbot system before a practitioner consultation will increase in the next few months [17]. The findings of the present study may support the development of information and prevention policies to avoid the misuse of AI by patients.

This study has several limitations, notably, the small number of patients, limiting the generalizability of the results; the use of ChatGPT-3.5 and not the latest version (GPT-4.0), the authors chose to evaluate the performance of the free version 3.5, more likely to be used by patients. The results of this study are therefore valid only for this free version, in french and without the ability to send clinical images or radiographs; more recent versions with better performance, would undoubtedly have obtained different results. ChatGPT's answers evolve with knowledge and its updates, so the evaluation of its performance having been carried out at a certain date is also susceptible to change over time. The evaluation was carried out by 3 surgeons, and was not double-blind, so they may not have been entirely objective, reducing the ChatGPT score. The primary strength of the study was the consideration of real clinical cases and not clinical theoretical vignettes. To the best of our knowledge, this study is the second one considering maxillofacial clinical cases in the assessment of ChatGPT performance.

At conclusion, ChatGPT may be a useful adjunctive clinical in daily practice tool to collect medical information. However, the current free version of ChatGPT appears less efficient in the proposition of the most probable primary diagnosis, the selection of the most adequate additional examination, and the proposition of pertinent and necessary therapeutic approaches. It would be interesting for future research to evaluate the latest ChatGPT version as well as new chatbots developed specifically for use in medical fields, which, if it proves to be more efficient, could be a useful tool in our daily practice as a source of information.

Informed consent

Patients consented to participate in the study.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Data availability statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Declaration of competing interest

The authors have no conflicts of interest.

CRediT authorship contribution statement

Mélissa Peters: Writing – original draft, Resources, Project administration, Methodology, Conceptualization. **Maxime Le Clercq:** Investigation, Data curation. **Antoine Yanni:** Software, Methodology, Investigation, Formal analysis, Data curation. **Xavier Vanden Eynden:** Methodology, Investigation, Formal analysis, Conceptualization.

Lalmand Martin: Methodology, Investigation, Formal analysis, Conceptualization. **Noémie Vanden Haute:** Investigation, Data curation. **Szonja Tancredi:** Investigation, Data curation. **Céline De Passe:** Investigation, Formal analysis. **Edward Boutremans:** Validation, Supervision, Project administration, Methodology. **Jerome Lechien:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Formal analysis, Conceptualization. **Didier**

Dequanter: Writing – review & editing, Validation, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Appendix A1. Clinical case features and trainee results

N	G	Age	Symptoms	History/medication	Clinical examination	Trainee consultation findings		
						Additional examinations	Diagnosis	Treatment
1	M	48	Bilateral preauricular swelling	Ear skin affection -> cremicort	Bilateral preauricular swelling	/	Bilateral viral parotitis	Antalgic I
2	M	29	Zygomatic pain L	6weeks after ORIF malar tripod fracture/ pregabaline, clonazepam	Mild swelling, heat, L malar tenderness	/	Post surgery infection	Amoxicillin/clavulanic acid, antalgic, antiseptic mouthwash
3	F	34	MO limitation	Orthodontic treatment (filed in 2020), joint clicking since	MO limited (2 cm), dental implants, painful masticatory muscles	/	TMD of muscular and articular origin (reversible dislocation of left articular disc)	Muscle and joint normalization physiotherapy
4	F	62	Paranasal mass L (since 1 month)	None	Mobile mass, 1cm	US, CT scan, MRI	Edema/nodule following facial vein thrombosis	Follow-up, removal if necessary
5	H	29	L jugal swelling (since 3 days)	Wisdom teeth extraction 10 years ago	L retromolar ulceronecrotic lesion	Panoramic radiography, biopsy, smear	Dental infection	Amoxicillin/clavulanic acid, analgesic I, mouthwash
6	H	62	R jugal swelling (since 5 days)	Dental extraction a few weeks ago	Jugal swelling, root residue 46, skin fistula	Biology, panoramic radiography, CT scan	Abscess with fistulization of dental origin	Amoxicillin/clavulanic acid, analgesic II, plan root extraction
7	F	46	R jugal swelling (since 6 days)	Collagen injection 11 days ago (cheeks, chin); allergy to mercury	R jugal swelling, cutaneous	/	Post-collagen injection skin infection	Amoxicillin/clavulanic acid, fusidic acid locally, analgesic I
8	M	20	Post-trauma facial tumefaction	Assault (blows to the face)	R supraorbital tumefaction, hypoesthesia V2 R, depressed R zygomatic arch	CT scan	R malar fracture	Soft diet, no valsava, analgesia I, Amoxicillin/clavulanic acid, antiseptic mouthwash. Reduction and osteosynthesis
9	F	86	Pain TMJ R	Knocked down by a scooter	Pain on palpation R TMJ, chin wound	CT scan	Mandibular fracture (bilateral intracapsular, displaced on right)	Conservator, soft diet, analgesic I, physiotherapy
10	F	50	L hemifacial hematoma	Knocked down on a scooter	Dermabrasion, hematoma L hemiface, hypoesthesia V1 L, V2 L	CT scan	Lateral and anterior wall R maxillary sinus fracture	Amoxicillin/clavulanic acid, no valsava ; re-evaluation
11	M	39	L jugal swelling (since 3 days) + pain L angle + fever/ chills Augmentin,	Trauma (blows to the face 5 days ago); ORIF L mandibular fracture 1 year ago	MO limited, L jugal swelling, painful	CT scan, biology	L peri-mandibular abscess	Amoxicillin/clavulanic acid analgesic I and II, drainage with drain
12	M	24	Dental displacement	Aggression (blows to the face)	Luxation 11,21,22,23,24	Ct scan	Alveolar bone fracture	Dental reduction and contention; amoxicillin/clavulanic acid, analgesic I and II, antiseptic mouthwash, soft diet
13	M	47	L TMJ swelling	Fall on scooter; dyslipidemia (atorvastatine), NSTEMI 2011	Chin wound, L TMJ swelling, MO limited, labial hematoma, mandibular deviation at OM	CT scan	Multi-fragmentary L subcondylar fracture	Intermaxillary fixation, amoxicillin/ clavulanic acid, analgesic I; wound suture
14	F	69	Chronic R jugal swelling	/	Induration at R stenson orifice, painful, no salivary flow at R stenson	CT scan	Chronic sialadenitis on lithiasis obstruction	Amoxicillin/clavulanic acid, sialogogue

(continued)

(Continued)

N	G	Age	Symptoms	History/medication	Clinical examination	Trainee consultation findings		
						Additional examinations	Diagnosis	Treatment
15	M	27	L jugal swelling (++ during meals)	Extraction of tooth 48 (2 days ago) / amoxicillin, paracetamol, antiseptic mouthwash	L jugal swelling, no salivary flow to L stenon, painful palpation of L parotid gland	/	Parotitis (partial obstruction of L stenon)	Sialogogues measures
16	M	60	L hemimandibular pain	Post tooth extraction infection 1 month ago / Clindamycin, H2O2 mouth wash	Purulent secretion, ulcerated lesion, bone involvement, induration floor of mouth	Smear, biopsy	Squamous cell carcinoma	Oncologic work-up and management
17	M	60	Oral mucosal lesion (1 month old)	Severe cough 2 weeks ago / multiple aphthosis 5 months ago / diabetes, hypertension / fluconazole, lidocaine and nystatin mouthwash/ amlodipine, metformin	Labial ulcerations, serosanguinous exudate, jugal leucoplasmia / single perineal lesion	Dermatological opinion	Erythema multiforme (post HSV or mycoplasma pneumonia)	Oral corticosteroid
18	M	39	L jugal swelling (since 7 days)	Amoxicilline 1 week, Amoxicillin/clavulanic acid since 1 day	L jugal swelling, L maxillary vestibular tumefaction, 27 painful on percussion	Panoramic radiography, CT scan	Abscess of dental origin	Drainage, amoxicillin/clavulanic acid, analgesic I, antiseptic mouthwash, dental opinion, surgical excision
19	M	37	Labial pain	Assault 3 days ago (blows to the face), labial wound sutured in emergency	Retroalveolar radiography	Intraoral Rx	Superinfection labial wound, contusion 21, 22	Amoxicillin/clavulanic acid, analgesic, antiseptic mouthwash
20	F	26	Intraoral purulent fluid	Epker osteotomy and genioplasty + PRF 6 days ago / Amoxicillin/clavulanic acid, analgesic I / anemia martial deficiency, malformative stenosis of the high pyelourethral junction	Limited MO, cloudy discharge from the neck of tooth 43	Smear, infectiologist opinion	Postoperative infection	Stop AB (according to infectiologist opinion for therapeutic window) then adjust antibiotic therapy
21	M	20	R mandibular tumefaction D (for 4 days)	15 g paracetamol yesterday	R paramandibular tumefaction, tender, MO limitation, vestibule tumefaction fourth quadrant; residual root 46, slightly indurated R submaxillary gland	Biology, CT scan	Facial cellulitis of dental origin (residual root 46)	Amoxicillin/clavulanic acid, NSAIDs, antiseptic mouthwash; monitoring for paracetamol intoxication
22	M	20	Hypersialorrhea	/	Correct salivary flow, no lesions	/	Hypersialorrhea due to anxiety	Patient reassurance
23	M	48	Intraoral swelling Panorex, dental check-up	Repeated abscess initially 38	Vestibular fistula 36, mobility 36, vital tooth	Panoramic radiography, check-up dental	Dental abscess	Amoxicillin/clavulanic acid, analgesic I, cyst curettage 36
24	F	9	L hemifacial swelling	Dental pain a few days ago, swimming in lake yesterday; bed fleas	L hemifacial tumefaction	Panoramic radiography, biology	Facial cellulitis of dental origin	Amoxicillin/clavulanic acid, pedodontic follow-up
25	F	65	Burning sensation in oral mucosa	Dental prosthesis	No lesions	Mycosis smear, allergy test, biology	Candidiasis or allergic reaction to dental materials	Stanford solution, analgesic
26	F	25	L jugal tumefaction (for 6 days) + purulent skin discharge	Amoxicillin for 6 days + NSAIDs Extraction 38 1 month ago	L peri-mandibular tumefaction, tender, indurated, skin fistula with purulent discharge, skin red and under tension, trismus	Panoramic radiography, biology, ct scan	Abscess on dental residue with fistulization to skin	Amoxicillin/clavulanic acid, analgesic, antiseptic skin gel, root extraction

(continued)

(Continued)

N	G	Age	Symptoms	History/medication	Clinical examination	Trainee consultation findings		
						Additional examinations	Diagnosis	Treatment
27	H	13	Pain L temporal and R under orbital	Aggression (blows to the face)	L temporal tumefaction, painful, R under orbital hematoma, painful	/	Contusion post-trauma	Cold pack, semi-seated position, analgesic I
28	H	26	Mandibular instability	Assault (blows to the face)/pulmonary tuberculosis 2016 and 2018; chronic psychiatric follow-up (insomnia, anxiety)/ mirtazapine, olanzapine, quetiapine, trazodone, escitalopram, nicotibine, rifadine, pyridoxine, clonazepam	Mandible unstable, mandibular fragment mobility between 43–44	CT scan	Bifocal mandibular fracture (R parasymphyseal, L displaced subcondylar)	Amoxicillin/clavulanic acid, analgesic, NSAID, antiseptic mouthwash, ligation, ORIF
29	M	26	L parotid swelling, + during meals	ORIF R parasymphyseal fracture and L subcondylar 5d ago with discharge required of patient after surgery /pulmonary tuberculosis 2016 and 2018; chronic psychiatric follow-up (insomnia, anxiety)/ mirtazapine, olanzapine, quetiapine, trazodone, escitalopram, nicotibine, rifadine, pyridoxine, clonazepam	L parotid swelling	Puncture	Postoperative seroma	AB, pressure dressing
30	M	3	Pre-maxillary pain	Trauma: fall + blow to the teeth	Tooth 61 very mobile, bleeding at the neck	/	Dental trauma with extrusion/luxation 61	Soft diet, analgesic I, follow-up in pedodontics
31	H	53	L paranasal swelling/ Patient	Rhinoplasty 2 months ago/clin-damycin/ atorvastatin	L paranasal induration, swelling, no inflammatory signs	/	Postoperative paranasal induration	Reassured, followed up with surgeon
32	F	6	L swelling hemifacial since 4 days, febrile peak with dental pain	Virosis with nasal congestion 7 days ago	L jugal swelling and L subpalpebral with redness; tooth 36 painful on percussion	Biolog, panoramic radiography	Ehtmoiditis	AB, analgesic I, ENT opinion
33	F	38	L jugal swelling for 4 days + jugal lesion for months/ Biopsy	/	L jugal and peri-mandibular swelling, budding lesion on L jugal mucosa, bleeding, necrotic areas	Biopsy	Suspicion carcinome épidermoïde Suspicion of epidermoid carcinoma	Based on biopsy results
34	F	37	R jugal swelling	Thyroidectomy/L-thyroxine	R jugal swelling, R vestibula swelling teeth 15–16, 15: pain on percussion; 16: dilapidated, pain on percussion, fistula	Panoramic radiography, bacteriological smear	Abscess of dental origin (tooth 15)	Drainage, rinse, amoxicillin/clavulanic acid, analgesic I, antiseptic mouthwash
35	H	32	Displaced teeth	Facial blows	L perimandibular tumefaction, peri-orbital ematoma, hypoesthesia L V3, MO limitation, space between 31–41, no L occlusion	Panoramic radiography, CT scan	Mandibular fracture (symphyseal and L subcondylar)	Amoxicillin/clavulanic acid, analgesic, NSAID, antiseptic mouthwash, ORIF

(continued)

(Continued)

N	G	Age	Symptoms	History/medication	Clinical examination	Trainee consultation findings		
						Additional examinations	Diagnosis	Treatment
36	H	28	Pain in tooth 47	Pulpectomy 2 days ago/antalgic level II	Pain palpation R mandibular border, hypoesthesia R V3, pain percussion 47	/	Acute apical parodontitis	Amoxicillin/clavulanic acid, analgesic II, urgent dental consultation
37	F	51	Pre-auricular pain for 10 years	Tooth extraction 10 years ago	R pre-auricular palpation pain, mandible deviates to the right at MO, refocuses at end of MO, masticatory muscle palpation pain, upper gingival ulceration	/	Joint and muscle TMD	Antalgic I, tramadol if necessary, diazepam at bedtime, physiotherapy; local flu-drocortisone/lidocaine
38	F	34	L jugal swelling for 7 days, pyrexia	/	Tooth 26 decayed, pain on percussion, vestibular swelling in the process of fistulation	Panoramic radiography	Abscess of dental origin	Incision and drainage, amoxicillin/clavulanic acid, dental care

G = gender (M = male, F = female); L = left; R = right; antalgic I, II or III = level I, II or III analgesic ; MO = mouth opening; ORIF = open reduction and internal fixation ; TMD = temporomandibular disorder; US = ultrasound;; V1, V2, V3 = nerve V1, V2, V3; TMJ = temporo-mandibular joint; AB = antibiotic; H2O2 = hydrogen peroxide

Appendix B1. ChatGPT findings regarding clinical cases

N	Practitioner	ChatGPT		
	Diagnosis	Additional examinations	Differential diagnosis - most likely diagnosis	Treatment according to clinical presentation
1	Bilateral viral parotitis	Biology, ultrasound, CT scan	Acute parotitis, bacterial parotitis , accessory salivary gland infection, middle ear infection	AB, hydration, rest, analgesic, ENT medical consultation
2	Post-operative infection	Blood culture, CT scan or MRI, bacteriological sampling	Post-operative infection , trigeminal neuralgia, reaction to clonazepam/pregabaline	AB, analgesic, surgical revision + drainage IN, close medical FU
3	Muscular and joint TMD	Panoramic radiography, CT scanner or MRI; biology	TMJ problem (disk disease or arthritis) , muscle disorder, coagulation disorder, dental or occlusion anomaly	Analgesics, NSAIDs, physiotherapy exercises, behavioral advice; orthodontics or prosthesis if occlusal problem; surgery if conservative treatment fails
4	Edema/nodule following facial vein thrombosis	Nasal endoscopy, CT scan or MRI, biopsy if nature of mass is unclear	Nasal polyp , chronic sinusitis, paranasal sinus tumor	AB if infection or sinusitis; excision of the mass if tumor or polyp, regular FU for monitoring
5	Dental infection	x-ray (intraoral or panoramic)	Dental infection , aphthous ulceration, mucocoele, salivary gland infection, malignant lesion	AB, dental care, analgesic I, dental FU
6	Abscess with fistulization of dental origin	X-ray (intraoral or panorex), US or Ct scanner or MRI, bacteriological sampling	Dental infection with abscess , salivary gland infection, skin infection, soft tissue tumor, cervical cellulitis	AB, dental care, dental FU
7	Post-collagen injection infection	Allergy test, US or CT scanner	Allergic reaction , infection, hematoma, autoimmune disorder	Antihistamine or corticosteroid (if allergy confirmed); dissolution of injected collagen IN; FU
8	R tripod malar fracture	CT scanner	Fracture of frontal bone , orbital floor, zygomatic arch ; V2 lesion, cranial trauma	Maxillofacial consultation, IN ORIF of fractures; analgesic, medical FU
9	Bilateral intracapsular fracture (displaced R fracture)	CT scanner	TMJ , mandible, zygomatic fracture; head trauma; cervical fracture	Maxillofacial consultation with IN surgery, analgesic, chin wound suture, FU
10	Lateral and anterior wall fracture of R maxillary sinus	CT scanner	Zygomatic fracture , mandible; cranial trauma, trigeminal nerve injury	Maxillofacial consultation, IN fracture surgery, dermabrasion treatment, analgesic, medical FU
11	Infection on osteosynthesis material, post trauma	CT scanner; blood culture	Post-traumatic infection , post-op inflammation, infection or reaction to osteosynthesis material; hematoma	Maxillofacial consultation, drainage and remove material IN; AB, analgesic, medical FU
12	Alveolar bone fracture	Dental x-ray, CT scanner	Maxillofracture (maxillary bone, tooth); dental injury , cranial trauma	Maxillofacial and dental consultation, IN ORIF of fractures, dentist: reduction, extraction or orthodontic treatment; IN wound suture; FU
13	Multi-fragmentary L subcondylar fracture	CT scanner, pre-op cardio assessment	TMJ fracture , mandible; skin lesion; bleeding disorder	Maxillofacial consultation, IN TMJ fracture ORIF, wound care, pre-op cardiac evaluation, analgesics, FU

(continued)

(Continued)

N	Practitioner	ChatGPT		
	Diagnosis	Additional examinations	Differential diagnosis - most likely diagnosis	Treatment according to clinical presentation
14	Hypothesis: Chronic sialadenitis due to stone obstruction	X-ray or CT scanner	Sialolithiasis , parotid infection, parotid tumor, stenon duct stenosis	Maxillofacial consultation, non-invasive stone removal (massage, irrigation), analgesic, AB IN, FU
15	Parotitis (partial L stenon obstruction)	US	Parotid infection , sialadenitis, post extraction infection	ENT or maxillofacial consultation, AB, antilagic, antiseptic mouthwash; FU
16	Squamous cell carcinoma	CT scanner	Mandibular osteomyelitis , dry socket, periodontal infection, bone tumor	Maxillofacial consultation, AB, irrigation and drainage, FU
17	Erythema multiforme	Biopsy, gynecological examination, biology	Oral candidiasis, oral lichen planus , oral herpes, recurrent oral aphthosis, perineal lesion	Dermato-oral and gynecological consultation, topical corticosteroid or immunosuppressant if lichen planus; FU
18	Abscess of dental origin	/	Dental infection (abscess) , facial cellulitis, parotid infection	Urgent dentistry consultation, AB, FU
19	Superinfection of labial wound, contusion of teeth 21–22	/	Labial wound infection , dental fracture	Cleaning, local antiseptic, dental care, AB, FU, wound care
20	Post-operative infection	X-ray, CT scanner	Post-operative infection , local inflammation	Consultation at his maxillo, AB to adapt IN, FU
21	Facial cellulitis of dental origin	Dental x-ray	Dental or periapical infection , submaxillary sialadenitis, mandibular soft tissue infection	Dental consultation, AB, analgesic, mouthwash, dental FU
22	Hypersialorrhea due to anxiety	Biology, ultrasound	Hypersialorrhea due to anxiety or medication, anxiety; swallowing problems	Advice, FU
23	Dental abscess	/	Periapical abscess , dental infection, periodontal infection	Extraction, AB, analgesic, monitoring
24	Cellulitis of dental origin	/	Facial cellulitis of dental origin , allergic reaction, dental infection, parasitic infection (bed flea)	Urgent dental consultation; IN: extraction, drainage, AB; analgesic, oral hygiene
25	Candidiasis or allergic reaction to dental material	Mycosis smear	Stomatodynia , xerostomia, allergy to dental materials, oral candidiasis	Dental consultation, prosthesis modification, topical anesthetic, hydration, oral hygiene
26	Abscess on dental residue with fistulization to the skin	/	Odontogenic infection (post-extraction 38) , cellulitis, skin abscess	AB, surgical drainage, extraction or endodontic treatment, analgesic, NSAIDs, FU
27	Post-traumatic contusion	/	Bone fracture (skull), facial contusion , suborbital hematoma	Analgesic, cold pack, monitoring, no aspirin
28	Bifocal mandibular fracture (R parasymphysis, L subcondylar displaced)	Panoramic radiography or CT scanner	Mandibular fracture , other facial fractures	Fracture reduction and immobilization, liquid or soft food, regular monitoring, analgesics, saline solution rinsing, psychiatric FU
29	Postoperative seroma	/	Parotid infection , post-operative complication (infection or inflammation)	AB, hydration, analgesic, follow-up, warm salt water mouth rinse, liquid or soft diet, psychiatric FU
30	Post-traumatic dental dislocation	Dental x-ray	Tooth fracture, tooth luxation	IN reduction and restraint, dental care, analgesics, NSAIDs, preventive advice
31	Post op paranasal induration	MRI, CT scanner or US; biopsy	Local infection, post-operative inflammatory reaction, tumor infiltrate	ENT or maxillo-facial consultation, IN surgery, FU
32	Ethmoiditis	/	Facial cellulitis , parotitis, dental abscess, sinusitis	AB (IV IN), analgesic, dental consultation, FU
33	Suspected squamous cell carcinoma	Biopsy, CT scan or MRI	Benign ulcer lesion, infection, malignant lesion	Depending on the stage: surgery, radiotherapy, chemotherapy, palliative care
34	Dental abscess on the 15)	Dental x-ray	Dental abscess on 16 parotitis, facial cellulitis	AB, endodontic treatment or extraction, analgesic, dental FU
35	Mandibular fracture (symphyseal and L subcondylar)	Panoramic radiography or CT scanner	Mandibular fracture , peri-orbital hematoma, traumatic V3 nerve injury	ORIF, temporary IMF, analgesic, FU
36	Acute apical parodontitis	Dental x-ray	Residual infection , nerve injury during pulpectomy, fracture 47	AB, endodontic advice, analgesic II
37	Joint and muscular TMD	x-ray, MRI	Trigeminal neuralgia, temporomandibular disorder , residual dental infection	Physiotherapy, orthosis, analgesics, NSAIDs, dietary advice
38	Abscess of dental origin	Panoramic radiography or retro alveolar x-ray	Dental infection , oral cellulitis or Ludwig's fasciitis	AB, dental care (extraction or endodontic treatment), analgesic

AB = antibiotic; ENT = ear, nose and throat specialist; IN: if necessary; FU = follow-up; MRI = magnetic resonance imaging; TMD = temporomandibular dysfunction; TMJ = temporomandibular joint; NSAID = nonsteroidal anti-inflammatory drug; US = ultra sound; R = right; L = left; V2 = maxillary nerve; IV = intravenous; V3 = mandibular nerve; ORIF = open reduction and internal fixation; IMF = intermaxillary fixation

References

- [1] 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):779–82 Nov.. doi: [10.1080/00016489.2023.2254809](https://doi.org/10.1080/00016489.2023.2254809).
- [2] Vaira LA, et al. Accuracy of ChatGPT-Generated Information on Head and Neck and Oromaxillofacial Surgery: a Multicenter Collaborative Analysis. *Otolaryngology–Head and Neck Surgery* 2024;170(6):1492–503 Jun.. doi: [10.1002/ohn.489](https://doi.org/10.1002/ohn.489).
- [3] Mago J, Sharma M. The Potential Usefulness of ChatGPT in Oral and Maxillofacial Radiology. *Cureus* 2023 Jul.. doi: [10.7759/cureus.42133](https://doi.org/10.7759/cureus.42133).
- [4] Wahlster W. Understanding computational dialogue understanding. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 2023;381(2251) Jul.. doi: [10.1098/rsta.2022.0049](https://doi.org/10.1098/rsta.2022.0049).
- [5] T.B. Brown et al., "Language Models are Few-Shot Learners," May 2020.
- [6] Domingos P. *The master algorithm: how the quest for the ultimate learning machine will remake our world.* New York.

- [7] Hill-Yardin EL, Hutchinson MR, Laycock R, Spencer SJ. A Chat(GPT) about the future of scientific publishing. *Brain Behav. Immun.* 2023;110:152–4 May. doi: [10.1016/j.bbi.2023.02.022](https://doi.org/10.1016/j.bbi.2023.02.022).
- [8] B. Mohammad et al., "The Pros and Cons of Using ChatGPT in Medical Education: a Scoping Review," 2023. [10.3233/SHIT230580](https://doi.org/10.3233/SHIT230580).
- [9] Alyasiri OM, Salman AM, Akhtom D, Salisu S. ChatGPT revisited: using ChatGPT-4 for finding references and editing language in medical scientific articles. *J. Stomatol. Oral Maxillofac. Surg.* 2024;101842 Mar.. doi: [10.1016/j.jormas.2024.101842](https://doi.org/10.1016/j.jormas.2024.101842).
- [10] Howard J. Artificial intelligence: implications for the future of work. *Am. J. Ind. Med.* 2019;62(11):917–26 Nov.. doi: [10.1002/ajim.23037](https://doi.org/10.1002/ajim.23037).
- [11] Lechien JR, Georgescu BM, Hans S, Chiesa-Estomba CM. ChatGPT performance in laryngology and head and neck surgery: a clinical case-series. *Eur Arch Oto-Rhino-Laryngol* 2024;281(1):319–33 Jan.. doi: [10.1007/s00405-023-08282-5](https://doi.org/10.1007/s00405-023-08282-5).
- [12] 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 Oto-Rhino-Laryngol* 2024;281(4):2063–79 Apr.. doi: [10.1007/s00405-023-08219-y](https://doi.org/10.1007/s00405-023-08219-y).
- [13] Rekman J, Hamstra SJ, Dudek N, Wood T, Seabrook C, Gofton W. A New Instrument for Assessing Resident Competence in Surgical Clinic: the Ottawa Clinic Assessment Tool. *J. Surg. Educ.* 2016;73(4):575–82 Jul.. doi: [10.1016/j.jsurg.2016.02.003](https://doi.org/10.1016/j.jsurg.2016.02.003).
- [14] Task Force for the Development of Student Clinical Performance Instruments, "The development and testing of APTA Clinical Performance Instruments. *Phys. Ther.* 2002;82(4):329–53 Apr..
- [15] Chen Y-Y, et al. Is the rating result reliable? A new approach to respond to a medical trainee's concerns about the reliability of Mini-CEX assessment. *J Formosan Medical Association* 2022;121(5):943–9 May. doi: [10.1016/j.jfma.2021.07.005](https://doi.org/10.1016/j.jfma.2021.07.005).
- [16] Chiesa-Estomba CM, et al. Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. *Eur Arch Oto-Rhino-Laryngol* 2024;281(4):2081–6 Apr.. doi: [10.1007/s00405-023-08104-8](https://doi.org/10.1007/s00405-023-08104-8).
- [17] A. A. and C.F.T.I. Jeffrey Millstein, "What can doctors and patients do with ChatGPT? | Expert Opinion," 2023.
- [18] Mira FA, et al. Chat GPT for the management of obstructive sleep apnea: do we have a polar star? *Eur Arch Oto-Rhino-Laryngol* 2024;281(4):2087–93 Apr.. doi: [10.1007/s00405-023-08270-9](https://doi.org/10.1007/s00405-023-08270-9).
- [19] Qu RW, Qureshi U, Petersen G, Lee SC. Diagnostic and Management Applications of ChatGPT in Structured Otolaryngology Clinical Scenarios. *OTo Open.* 2023;7(3) Jul.. doi: [10.1002/oto2.67](https://doi.org/10.1002/oto2.67).
- [20] Lechien JR, Chiesa-Estomba C-M, Baudouin R, Hans S. Accuracy of ChatGPT in head and neck oncological board decisions: preliminary findings. *Eur Arch Oto-Rhino-Laryngol* 2024;281(4):2105–14 Apr.. doi: [10.1007/s00405-023-08326-w](https://doi.org/10.1007/s00405-023-08326-w).
- [21] Frosolini A, et al. Assessing the accuracy of ChatGPT references in head and neck and ENT disciplines. *Eur Arch Oto-Rhino-Laryngol* 2023;280(11):5129–33 Nov.. doi: [10.1007/s00405-023-08205-4](https://doi.org/10.1007/s00405-023-08205-4).
- [22] Hoch CC, 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 Oto-Rhino-Laryngol* 2023;280(9):4271–8 Sep.. doi: [10.1007/s00405-023-08051-4](https://doi.org/10.1007/s00405-023-08051-4).