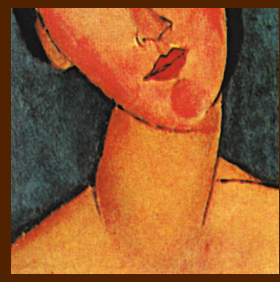
OTORHINOLARYNGOLOGY

PUBBLICAZIONE PERIODICA TRIMESTRALE - POSTE ITALIANE S.P.A. - SPED. IN A. P. D.L. 353/2003 (CONV. IN L. 27/02/2004 N° 46) ART. 1, COMMA 1, DCB/CN - ISSN 2724-6302 TAXE PERÇUE



QUARTERLY JOURNAL ON
OTORHINOLARYNGOLOGY,
AUDIOLOGY
PHONATRICS, HEAD AND
NECK SURGERY
MAXILLO-FACIAL SURGERY
PLASTIC RECONSTRUCTIVE
SURGERY
OTONEUROSURGERY

VOL.75 No.4
DECEMBER 2025

EDIZIONI MINERVA MEDICA

Otorhinolaryngology 2025 December;75(4):120-6
DOI: 10.23736/S2724-6302.25.02603-9

ORIGINAL ARTICLE

Validation of the QAMAI tool in Italian for the evaluation AI-generated health information in head and neck surgery

Luigi A. VAIRA [1] *, Giacomo DE RIU [1], Giovanni SALZANO [2], Carlos M. CHIESA-ESTOMBA [3], Giuseppe CONSORTI [4], Giulio CIRIGNACO [4], Fabio MAGLITTO [2], Antonino MANIACI [5], Miguel MAYO-YANEZ [6], Marzia PETROCELLI [7], Alberto M. SAIBENE [8], Stefania TROISE [2], Andrea DE VITO [9], Sholem HACK [10], Francesco LAGANÀ [11], Bernardo BIANCHI [11], Paolo BOSCOLO-RIZZO [12], Jerome R. LECHIEN [13, 14]

[1] Operative Unit of Maxillo-Facial Surgery, Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy; [2]Head and Neck Section, Department of Neurosciences, Reproductive and Odontostomatological Sciences, Federico II University of Naples, Naples, Italy; [3]Department of Otorhinolaryngology-Head and Neck Surgery, Hospital Universitario Donostia, San Sebastian, Spain; [4]Division of Maxillofacial Surgery, Department of Neurological Sciences, Marche University Hospitals-Umberto I, Ancona, Italy; [5]Department of Medicine and Surgery, University of Enna Kore, Enna, Italy; [6]Otorhinolaryngology, Head and Neck Surgery Department, Complexo Hospitalario Universitario A Coruña (CHUAC), A Coruña, Galicia, Spain; [7]Maxillofacial Surgery Operative Unit, Bellaria and Maggiore Hospital, Bologna, Italy; [8]Otolaryngology Unit, Santi Paolo e Carlo Hospital, Department of Health Sciences, University of Milan, Milan, Italy; [9]Operative Unit of Infectious and Tropical Diseases, Department of Medicine, Surgery and Pharmacy, University of Sassari, Sassari, Italy; [10]City St. Georges University London School of Medicine, Program Delivered by University of Nicosia at the Chaim Sheba Medical Center, Ramat Gan, Israel; [11]Maxillofacial Surgery Unit, IRCCS San Martino Polyclinic Hospital, Genoa, Italy; [12]Department of Medical, Surgical and Health Sciences, Section of Otolaryngology, University of Trieste, Trieste, Italy; [13]Department of Surgery, Mons School of Medicine, UMONS Research Institute for Health Sciences and Technology, University of Mons (UMons), Mons, Belgium; [14]Department of Otolaryngology-Head Neck Surgery, Elsan Hospital, Paris, France

*Corresponding author: Luigi A. Vaira, Operative Unit of Maxillo-Facial Surgery, Department of Medicine, Surgery and Pharmacy, viale San Pietro 43B, 07100 Sassari, Italy. E-mail: lavaira@uniss.it

ABSTRACT

BACKGROUND: This study aimed to validate the Italian version of the Quality Assessment of Medical Artificial Intelligence (IT-QAMAI) tool, designed to evaluate the reliability of AI-generated health information in the context of head and neck surgery.

METHODS: The IT-QAMAI tool was adapted from the original English version and involved a rigorous translation and back-translation process. The validation involved 18 researchers from 13 centers across Europe, assessing 24 AI-generated responses categorized into clinical scenarios, theoretical questions, and patient inquiries. The tool's reliability was measured using Cronbach's alpha for internal consistency, the Intraclass Correlation Coefficient (ICC) for inter-rater reliability, and Pearson's correlation for test-retest reliability.

RESULTS: The IT-QAMAI demonstrated high internal consistency (Cronbach's alpha = 0.850) and good inter-rater reliability (ICC=0.750). Test-retest reliability was strong (r_s=0.887). Significant differences were found in the quality of AI-generated responses across different question types.

CONCLUSIONS: The IT-QAMAI tool is a reliable and valid instrument for assessing the quality of AI-generated health information in Italian, with significant implications for its use in clinical practice and research in head and neck surgery.

(*Cite this article as:* Vaira LA, De Riu G, Salzano G, Chiesa-Estomba CM, Consorti G, Cirignaco G, *et al*. Validation of the QAMAI tool in Italian for the evaluation AI-generated health information in head and neck surgery. Otorhinolaryngology 2025;75:120-6. DOI: 10.23736/S2724-6302.25.02603-9)

KEY WORDS: Artificial intelligence; Health information systems; Otolaryngology; Oral surgery; Quality control.

The rapid advancement of Artificial Intelligence (AI) technologies has significantly influenced various sectors, with healthcare being one of the most impacted.[1-3] AI platforms, particularly natural language processing models like ChatGPT, have the potential to revolutionize the dissemination of health information by providing timely, personalized, and easily accessible content to a broad audience.[4, 5] However, the use of AI in healthcare also presents substantial risks, particularly concerning the accuracy, reliability, and potential misuse of the information provided.[6] These risks highlight the urgent need for validated tools to evaluate the quality of health information generated by AI systems.[7]

Despite this pressing need, the development of validated tools for assessing AI-generated health information is still in its infancy. Currently, only two tools have been proposed: the Quality Assessment of Medical Artificial Intelligence (QAMAI) tool,[8] which focuses on evaluating the quality of health information provided by AI platforms, and the Artificial Intelligence for Pathway Integrity (AIPI) score[9] which assesses the quality of diagnostic-therapeutic pathways suggested by AI systems. Unfortunately, the most studies published to date rely on non-validated systems, which can introduce significant issues, such as inconsistent assessments, biased results, and unreliable conclusions.[10-15] These shortcomings underscore the critical importance of developing and validating robust evaluation tools to ensure that AI-generated health information is accurate, reliable, and applicable across different linguistic and cultural contexts. While the QAMAI tool has been validated in English, it is important to recognize the necessity of validating such tools in the languages in which they are applied. The accuracy and relevance of information can be influenced by linguistic and cultural contexts, making localized validation essential for ensuring the tool's effectiveness.[16, 17]

The objective of this study is to propose and validate the Italian version of the Quality Assessment of Medical Artificial Intelligence (IT-QAMAI) tool. Given the increasing use of AI platforms like ChatGPT by Italian-speaking users, it is crucial to have a reliable tool for assessing the quality of health information in Italian.

Materials and methods

This study was conducted as part of a research initiative established in February 2023, with the primary aim of exploring the potential applications, reliability, and safety of AI in the healthcare field, with a particular focus on head and neck surgery. The research group was composed of otolaryngologists, maxillofacial surgeons, and dentists from various centers around the world.

For this specific study, 18 researchers from 13 centers across Italy, France, Belgium and Spain were involved. The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval from an ethics committee was not required, as the study did not involve human patients or animals.

Adaptation of the instrument

The adaptation of the Italian version of the QAMAI tool was based on the original English version (Supplementary Digital Material 1: Supplementary Table I) and followed established guidelines for the translation and cultural adaptation of questionnaires to different languages.[17, 18] The process involved several steps to ensure accuracy and cultural relevance.

Initially, two independent bilingual translators, both native Italian speakers, translated the questionnaire from English to Italian. One translator was a head and neck surgeon with over 20 years of experience, while the other was a layperson with no prior knowledge of the questionnaire's purpose. Discrepancies between the translations were resolved through discussion between the translators, with a third independent translator available to mediate if necessary.

The translated Italian version was then back-translated into English by two different bilingual translators, both native English speakers who were not familiar with the content or purpose of the questionnaire. This back-translation was used to check for accuracy and to identify any potential misunderstandings or discrepancies in the initial translation.

Following the translation and back-translation, the Italian version of the questionnaire was reviewed by a panel of experts, including native Italian-speaking head and neck surgeons, a computer engineer with expertise in AI, a bioethics specialist, and a communications engineer with a focus on health communication. The panel carefully reviewed each item, refining the translation to ensure clarity and accuracy. This resulted in a prefinal version of the tool.

The prefinal version was tested with a sample of four respondents. These respondents were asked to use the tool to evaluate a set of responses and to describe their understanding of each question-and-answer option. The expert panel then analyzed these responses to ensure that the questions and answers were correctly interpreted and free from misunderstandings. Based on this feedback, the

TABLE I.—*The Italian version of the Quality Analysis of Medical Artificial Intelligence tool (IT-QAMAI).*

Item	1 Fortemente in disaccordo	2 In disaccordo	3 Neutro	4 D'accordo	5 Fortemente d'accordo
Accuratezza: Le informazioni fornite sono accurate e aggiornate.					
Chiarezza: La risposta è chiara e comprensibile in termini di linguaggio e terminologia scientifica.					
Rilevanza: Le informazioni fornite sono pertinenti e rispondono direttamente alla domanda posta.					
Completezza: La risposta copre adeguatamente tutti gli aspetti della domanda e fornisce informazioni sufficienti, inclusi eventuali elementi di incertezza.					
Fornitura di fonti e riferimenti: La risposta fornisce fonti e riferimenti affidabili a supporto delle informazioni sanitarie presentate.					
Utilità: La risposta è utile per soddisfare le esigenze informative dell'utente in ambito sanitario.					

panel reached a new consensus on each item, resulting in the final version of the IT-QAMAI (Table I).

Data collection

A team of three head and neck surgeons prepared a set of 24 questions to be submitted to the AI, categorized into three subgroups of eight questions each: clinical scenarios, theoretical questions, and patient inquiries. These questions covered three key areas of head and neck surgery: oncologic surgery, trauma surgery, and nasal-sinus surgery. The questions were reviewed by the research group and revised as needed to correct any errors or address areas of uncertainty, until a full consensus was reached. The entire set of questions is provided in Supplementary Digital Material 2 (Supplementary Table II).

Each question was individually input by a single researcher into the ChatGPT version 4 chatbot[19] on May 3, 2024. To ensure consistency and accuracy, the researcher used a new browser window in incognito mode for each entry. The AI was instructed to provide the most complete and detailed responses possible, including the bibliographic sources from which the information was derived. Different SMART prompts[20] were employed for each category of questions to optimize the quality of the AI's responses (Supplementary Table II).

The responses generated by the AI were then collected by the researcher and provided to a panel of eight reviewers for blind evaluation (Supplementary Table II). This panel consisted of head and neck surgeons with over 10 years of experience. Each reviewer independently assessed the responses, resulting in eight evaluations for each answer, which were then used in the validation process. The evaluation was conducted using the IT-QAMAI tool.

The QAMAI evaluates various aspects such as accuracy, clarity, relevance, completeness, provision of sources, and overall usefulness. The scoring system is based on a Likert scale, where reviewers rate each item from 1 (strongly disagree) to 5 (strongly agree), resulting in a cumulative score, ranging from 5 to 30, that reflects the overall quality of the AI's responses.

To assess test-retest reliability, the evaluation was repeated 10 days after the initial assessment by the same panel of reviewers.

Statistical analysis

Statistical analyses were conducted using Jamovi version 2.3.18.0, a free and open-source statistical software available online at www.jamovi.org. Categorical variables are presented as numerals and percentages of the total, while descriptive statistics for quantitative variables are reported as either the median with interquartile range (IQR) or mean with standard deviation (SD). Differences in the IT-QAMAI scores across the three categories of questions were evaluated using a one-way ANOVA. If significant differences were identified, the Dwass-Steel-Crichlow-Fligner test was employed for post-hoc analysis.

Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) test were used to assess the adequacy of the sampling, with a KMO value greater than 0.6 indicating sufficient sampling.[21]

In the original version of the QAMAI tool, exploratory factor analysis revealed that the tool had a unidimensional nature, with a single factor encompassing all the items.[8] However, due to the lack of a gold standard for comparison, we conducted a new exploratory factor analysis to revalidate the tool from scratch. This analysis aimed to uncover the interrelationships between clusters of items and to determine the number of factors assessed by the ques-

tionnaire. The analysis employed a minimum residual extraction method with promax rotation, using a cut-off point of 0.40 and Kaiser's criterion of eigenvalues greater than 1. Following the identification of factors through exploratory factor analysis, a confirmatory factor analysis was performed to verify the fit of the data to the identified theoretical model. The model's goodness-of-fit was evaluated using the root mean square error of approximation (RMSEA) and the Comparative Fit Index (CFI), with RMSEA values below 0.05 and CFI values above 0.95 considered indicative of a good model fit.[22]

Internal consistency of the tool was assessed using Cronbach's alpha to determine whether the items were inter-correlated and consistent with the underlying construct of the tool. A Cronbach's alpha value of 0.70 or higher was considered to indicate adequate internal consistency.[23]

Inter-rater reliability was evaluated by comparing the ratings provided by different reviewers for each question using the intraclass correlation coefficient (ICC). The ICC values were interpreted as follows: ICC below 0.5 indicated poor reliability, ICC between 0.5 and 0.75 indicated moderate reliability, ICC between 0.75 and 0.9 indicated good reliability, and ICC above 0.9 indicated excellent reliability.[23]

Finally, the test-retest reliability between the two sets of evaluations conducted by researchers, spaced 10 days apart, was assessed using Pearson's correlation coefficient. A P value of less than 0.05 was considered to indicate statistical significance for all tests.

Results

A total of 184 evaluations were provided by the 8 reviewers across the 24 questions, resulting in an overall median QAMAI score of 22 [IQR 20-24]. Significant differences in the quality of responses were observed among the three different types of questions. Specifically, patient inquiries received significantly higher QAMAI scores compared to clinical scenarios (Figure 1). However, the differences between patient inquiries and theoretical questions, as well as between theoretical questions and clinical scenarios, were not statistically significant (Figure 1).

Of the 184 quality assessments collected, 8 were rated as excellent (QAMAI score 30), 53 as very good (QAMAI score 24-29), 94 as good (QAMAI score 18-23), 28 as fair (QAMAI score 12-17), and 1 as poor (QAMAI score 6-11).

The adequacy of the sample was assessed using Bartlett's test of sphericity, which yielded a significant result (χ^2=542; P<0.001), indicating that the data were suitable for factor analysis. Following this, the KMO test was performed, which produced an overall value of 0.862 (range: 0.823-0.912). This result indicates excellent sampling adequacy.

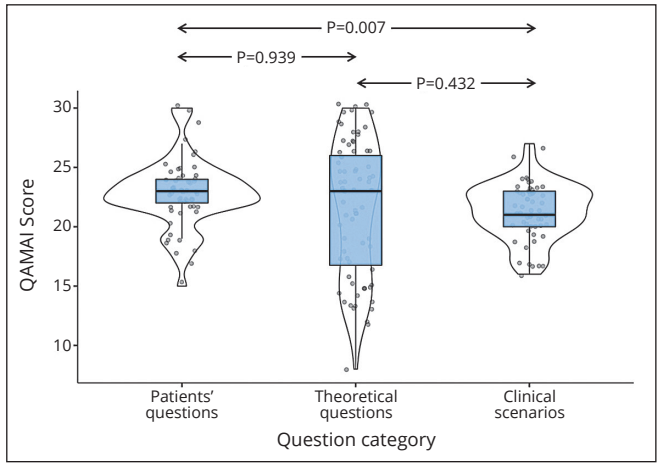


Figure 1.—Distribution of IT-QAMAI Scores across different question categories.

The exploratory factor analysis conducted on the IT-QAMAI tool indicated a unidimensional structure, consistent with the findings from the original version of the tool (Figure 2). This single factor accounted for 54.8% of the total variance, with factor loadings for individual items ranging from 0.399 to 0.887. The model fit was further evaluated using the RMSEA, which yielded a value of 0.0852 (90% CI: 0.0377 - 0.134), and the Comparative Fit Index CFI with a value of 0.962.

The internal consistency of the IT-QAMAI tool was assessed using Cronbach's alpha. The analysis yielded a Cronbach's alpha value of 0.850, indicating a high level

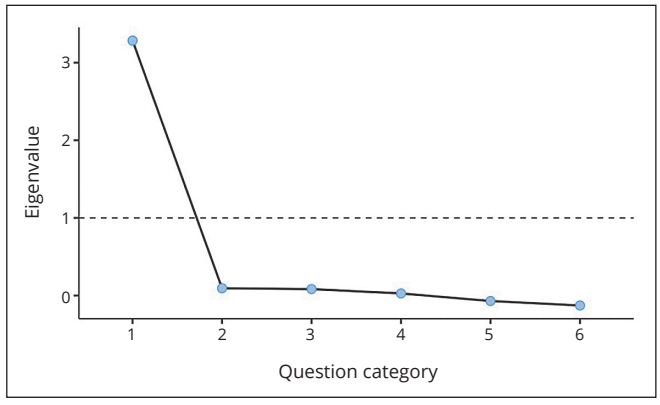


Figure 2.—Scree plot for exploratory factor analysis.

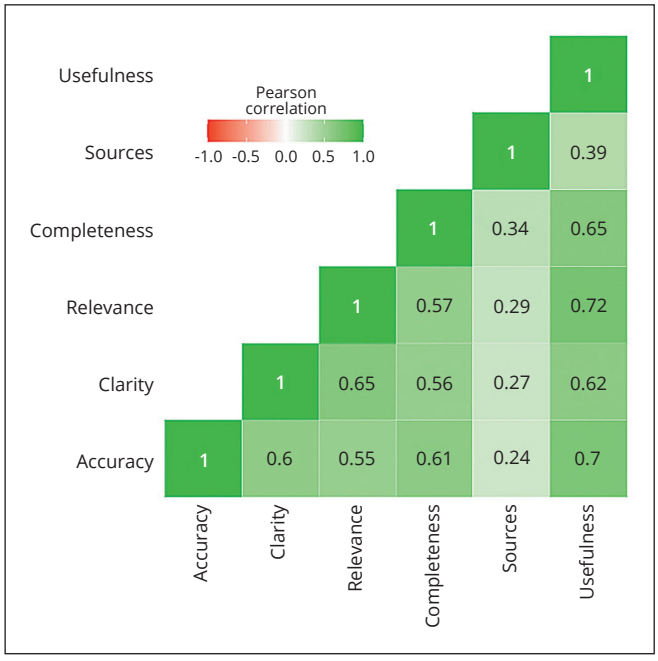


Figure 3.—Correlation matrix of IT-QAMAI Tool items.

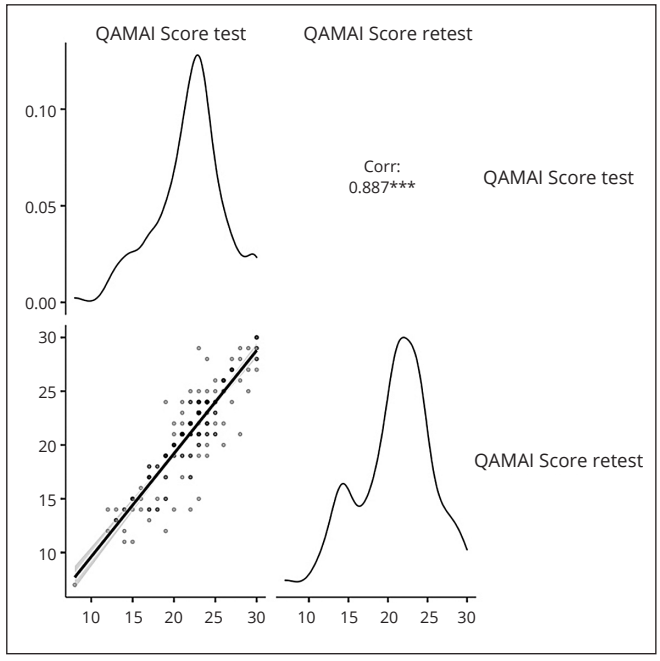


Figure 4.—Test-retest reliability of IT-QAMAI Scores.

of internal consistency (Figure 3). This result suggests that the items within the IT-QAMAI tool are well-correlated and consistently measure the same underlying construct.

The inter-rater reliability of the IT-QAMAI tool was evaluated using the ICC. The analysis produced an ICC value of 0.750 (95% CI: 0.639-0.850; P<0.001), indicating a good level of agreement among the eight reviewers.

Finally, the test-retest reliability of the IT-QAMAI tool was assessed by comparing the QAMAI scores from two evaluations conducted 10 days apart. The Pearson correlation coefficient for the test-retest scores was 0.887 (95% CI: 0.852 to 0.914; P<0.001), indicating a strong positive and statistically significant correlation (Figure 4).

Discussion

The findings of this study provide evidence supporting the reliability and validity of the IT-QAMAI tool. This study was undertaken to address the urgent need for validated tools that can assess the quality of AI-generated health information, particularly in non-English-speaking contexts. As AI continues to become more integrated into healthcare, ensuring that the information it provides is accurate and reliable is crucial for both patient safety and the integrity of healthcare systems.[24]

The IT-QAMAI tool demonstrated a unidimensional structure, consistent with the original English version, confirming that it effectively measures the quality of AI-generated health information across various dimensions such as accuracy, clarity, relevance, and completeness. The factor analysis results, showing that a single factor explained 54.8% of the variance, are comparable to the findings in the original tool, further validating its applicability in the Italian context. These results align with other studies that emphasize the importance of having culturally and linguistically adapted tools for evaluating the quality of the information.[16, 17]

The internal consistency of the IT-QAMAI tool, as indicated by a Cronbach's alpha of 0.850, suggests that the tool is highly reliable. This level of consistency is crucial for ensuring that the tool provides dependable results across different contexts and user groups.

The study also demonstrated good inter-rater reliability, with an ICC value of 0.750. This indicates that the tool produces consistent results across different evaluators, which is critical for ensuring that assessments of AI-generated health information are objective and not subject to individual bias. The high ICC value found in this study is comparable to other well-validated tools used in healthcare settings,[9, 25] highlighting the IT-QAMAI tool's reliability.

Additionally, the test-retest reliability assessment showed a strong Pearson correlation coefficient of 0.887

between the two evaluations conducted 10 days apart. This indicates that the tool provides stable and consistent results over time, which is essential for longitudinal studies and repeated assessments of AI-generated content.

The validation of the IT-QAMAI tool has significant implications for both clinical practice and research. As AI platforms become more prevalent in healthcare, tools like the IT-QAMAI will be indispensable for evaluating the quality of the information these platforms provide. This tool can help healthcare professionals, researchers, and policymakers ensure that AI-generated content meets high standards of accuracy and reliability, ultimately contributing to better patient outcomes and more informed decision-making.

Furthermore, the IT-QAMAI tool's validation in Italian opens the door for its adaptation and validation in other languages and cultural contexts. As AI technology continues to spread globally, the need for reliable, culturally sensitive tools to assess AI-generated content will only grow. The methodology used in this study can serve as a model for similar adaptations in other languages, ensuring that AI assessment tools are relevant and effective across diverse populations.

Limitations of the study

Several limitations should be acknowledged. First, the study focused exclusively on head and neck surgery, which may limit the generalizability of the findings to other medical fields. Future research should explore the applicability of the IT-QAMAI tool in other areas of healthcare to determine whether its reliability and validity hold across different specialties.

Additionally, the sample size, while adequate for this validation study, could be expanded in future research to include a more diverse range of respondents and AI-generated content. Larger, more diverse samples would provide more robust data and help to further validate the tool's applicability in various contexts.

Finally, as AI technology continues to evolve, so too must the tools used to evaluate its outputs. Ongoing research and periodic updates to the IT-QAMAI tool will be necessary to ensure that it remains relevant and effective in assessing the quality of AI-generated health information.

Conclusions

In conclusion, the IT-QAMAI tool demonstrated strong reliability and validity. This tool can provide a valuable resource for Italian researcher and professionals for evaluating the quality of AI-generated health information, ensuring that such content is accurate, reliable, and suitable for use in healthcare. As AI continues to play an increasingly prominent role in healthcare, tools like IT-QAMAI will be essential for safeguarding the quality of information and supporting informed decision-making in clinical practice. Future research should continue to explore and refine these tools, ensuring that they keep pace with the rapid advancements in AI technology and its applications in healthcare.

References

1. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018;2:719–31.

2. 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.

3. 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;170:1492–503.

4. Cheng K, Sun Z, He Y, Gu S, Wu H. The potential impact of ChatGPT/GPT-4 on surgery: will it topple the profession of surgeons? Int J Surg 2023;109:1545–7.

5. Radulesco T, Saibene AM, Michel J, Vaira LA, Lechien JR. ChatGPT-4 performance in rhinology: A clinical case series. Int Forum Allergy Rhinol 2024;14:1123–30.

6. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023;6:1169595.

7. Lavalle S, Lechien JR, Chiesa-Estomba C, Parisi FM, Maniaci A. Evaluating AI in patient education: the need for a validated performance assessment tool. Am J Otolaryngol 2024;45:104442.

8. Vaira LA, Lechien JR, Abbate V, Allevi F, Audino G, Beltramini GA, *et al.* 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 2024;281:6123–31.

9. Lechien JR, Maniaci A, Gengler I, Hans S, Chiesa-Estomba CM, Vaira LA. Validity and reliability of an instrument evaluating the performance of intelligent chatbot: the Artificial Intelligence Performance Instrument (AIPI). Eur Arch Otorhinolaryngol 2024;281:2063–79.

10. Deiana G, Dettori M, Arghittu A, Azara A, Gabutti G, Castiglia P. Artificial Intelligence and Public Health: Evaluating ChatGPT Responses to Vaccination Myths and Misconceptions. Vaccines (Basel) 2023;11:1217.

11. Chiesa-Estomba CM, Lechien JR, Vaira LA, Brunet A, Cammaroto G, Mayo-Yanez M, *et al.* Exploring the potential of Chat-GPT as a supportive tool for sialendoscopy clinical decision making and patient information support. Eur Arch Otorhinolaryngol 2024;281:2081–6.

12. Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, *et al.* Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model. Res Sq 2023;rs.3.rs-2566942. [Preprint]

13. Lechien JR, Naunheim MR, Maniaci A, Radulesco T, Saibene AM, Chiesa-Estomba CM, *et al.* Performance and Consistency of ChatGPT-4 Versus Otolaryngologists: A Clinical Case Series. Otolaryngol Head Neck Surg 2024;170:1519–26.

14. Mayo Yanez M, Gonzalez-Torres L, Saibene AM, Allevi F, Vaira LA, Maniaci A, *et al.* Application of ChatGPT as a support tool in the diagnosis and management of acute bacterial tonsillitis. Health Technol 2024;14:773–9.

15. Saibene AM, Allevi F, Calvo-Henriquez C, Maniaci A, Mayo-Yáñez M, Paderno A, *et al.* Reliability of large language models in managing odontogenic sinusitis clinical scenarios: a preliminary multidisciplinary evaluation. Eur Arch Otorhinolaryngol 2024;281:1835–41.

16. Hawkins M, Cheng C, Elsworth GR, Osborne RH. Translation method is validity evidence for construct equivalence: analysis of secondary data routinely collected during translations of the Health Literacy Questionnaire (HLQ). BMC Med Res Methodol 2020;20:130.

17. Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. Saudi J Anaesth 2017;11(Suppl 1):S80–9.

18. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011;17:268–74.

19. ChatGPT. 4; 2025 [Internet]. Available from: https://openai.com/blog/chatgpt [cited 2025, Dec 10].

20. Vaira LA, Lechien JR, Abbate V, Gabriele G, Frosolini A, De Vito A, *et al.* Enhancing AI chatobot responses in healthcare: the SMART prompt structure in head and neck surgery. OTO Open 2025;9:e70075.

21. Dziuban CD, Shirkey EC. When is a correlation matrix appropriate for factor analysis? Some decision rules. Psychol Bull 1974;81:358–61.

22. Wolf MG, McNeish D. Dynamic: an R package for deriving dynamic fit index cutoffs for factor analysis. Multivariate Behav Res 2023;58:189–94.

23. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63.

24. Yu P, Xu H, Hu X, Deng C. Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration. Healthcare (Basel) 2023;11:2776.

25. Vaira LA, Sergnese S, Salzano G, Maglitto F, Arena A, Carraturo E, *et al.* Are YouTube Videos a Useful and Reliable Source of Information for Patients with Temporomandibular Joint Disorders? J Clin Med 2023;12:817.

Conflicts of interest
The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Funding
This work has been developed within the framework of the project e.INS- Ecosystem of Innovation for Next Generation Sardinia (cod. ECS 00000038) funded by the Italian Ministry for Research and Education (MUR) under the National Recovery and Resilience Plan (NRRP) - MISSION 4 COMPONENT 2, "From research to business" INVESTMENT 1.5, "Creation and strengthening of Ecosystems of innovation" and construction of "Territorial R&D Leaders," CUP J83C21000320007.

Authors' contributions
Luigi A. Vaira and Giacomo De Riu have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Conceptualization and methodology: Luigi A. Vaira, Jerome R. Lechien. Acquisition, analysis and interpretation of the data: Giovanni Salzano, Giuseppe Consorti, Giulio Cirignaco, Fabio Maglitto, Marzia Petrocelli, Stefania Troise, Alberto M. Saibene, Francesco Laganà. Statistical analysis: Antonino Maniaci, Bernardo Bianchi. Critical revision of the literature: Jerome R. Lechien, Carlos M. Chiesa-Estomba, Miguel Mayo-Yanez, Sholem Hack, Andrea De Vito, Paolo Boscolo-Rizzo. Provision of resources: Giacomo De Riu. Original draft preparation: Luigi A. Vaira, Giacomo De Riu. Critical revision of the manuscript for important intellectual content: Giovanni Salzano, Carlos M. Chiesa-Estomba, Giuseppe Consorti, Giulio Cirignaco, Fabio Maglitto, Antonino Maniaci, Miguel Mayo-Yanez, Marzia Petrocelli, Alberto M. Saibene, Stefania Troise, Andrea De Vito, Sholem Hack, Francesco Laganà, Bernardo Bianchi, Paolo Boscolo-Rizzo, Jerome R. Lechien. Supervision: Jerome R. Lechien. All the authors read and approved the final version of the manuscript.

Acknowledgements
The authors would like to acknowledge the assistance of ChatGPT-4o (OpenAI, San Francisco, CA, USA) for proofreading and language refinement of this manuscript. The use of AI-assisted tools was limited to grammatical and stylistic suggestions, and all intellectual content remains the sole responsibility of the authors.

History
Manuscript accepted: November 5, 2025. - Manuscript received: October 9, 2025.

Supplementary data
For supplementary materials, please see the HTML version of this article at www.minervamedica.it