



Published in final edited form as:

Laryngoscope. 2025 October ; 135(10): 3579–3587. doi:10.1002/lary.32202.

An Evaluation of Current Trends in AI-Generated Text in Otolaryngology Publications

Rachel B. Kutler, BA¹, Sean A. Setzen, MD¹, Samantha Tsai¹, Anaïs Rameau, MD MSc MPhil MS FACS¹

¹.) Sean Parker Institute for the Voice, Department of Otolaryngology-Head and Neck Surgery, Weill Cornell Medical College, New York, New York

Abstract

Objectives: Since the release of ChatGPT-4 in March 2023, large language models (LLM) application in biomedical manuscript production has been widespread. GPT-modified text detectors, such as GPTzero, lack sensitivity and reliability and do not quantify the amount of AI-generated text. However, recent work has identified certain adjectives more frequently used by LLMs that can help identify and quantify LLM-modified text. The aim of this study is to utilize these adjectives to identify LLM-generated text in otolaryngology publications.

Study Design: Meta-research

Methods: Twenty-five otolaryngology journals were studied between November 2022 and July 2024, encompassing 8,751 published works. Articles from countries where ChatGPT-4 is not available were removed, yielding 7,702 articles for study inclusion. These publications were analyzed using a Python script to determine the frequency of the top 100 adjectives disproportionately generated by ChatGPT-4.

Results: A significant increase in the frequency of adjectives associated with GPT use was observed from November 2023 to July 2024 across all journals ($p < 0.001$), with a significant difference before and after the release of ChatGPT in March 2023. Journals with higher impact factors had significantly lower usage of GPT-associated adjectives than those with lower impact factors ($p < 0.001$). There was no significant difference in GPT-associated adjective use by first authors with a doctoral degree versus those without. Publications by authors from English-speaking countries demonstrated a significantly more frequent use of LLM-associated adjectives ($p < 0.001$).

Corresponding Author: Anaïs Rameau, MD, MPhil, MS, Sean Parker Institute for the Voice, Department of Otolaryngology-Head and Neck Surgery, Weill Cornell Medical College, 240 E 59th Street, New York, NY 10022, anr2783@med.cornell.edu.

Author Contributions (conception, design, data collection, literature review, drafting, data analysis, draft revision, etc.)

Rachel B. Kutler: data collection, drafting, data analysis

Sean A. Setzen: data collection, drafting, data analysis

Samantha Tsai: data collection

Anaïs Rameau: conception, design, manuscript revisions

Level of Evidence: N/A

Conflict of Interest: Anaïs Rameau owns equity of Perceptron Health Inc. She is an advisor for Sound Health Inc. and Pentax Medical.

Conclusions: This study suggests that ChatGPT use in otolaryngology manuscript production has significantly increased since the release of ChatGPT-4. Future research should be aimed at further characterizing the landscape of AI-generated text in otolaryngology and developing tools that encourage authors' transparency regarding use of LLMs.

Graphical Abstract

This study evaluates the prevalence of AI-generated text in otolaryngology publications following the release of ChatGPT-4 in March 2023 by analyzing the frequency of GPT-associated adjectives in 6,325 articles from 25 journals. Findings indicate a significant increase in AI-generated text over time, with lower impact factor journals and authors from English-speaking countries exhibiting higher usage. These results highlight the growing influence of LLMs in medical writing and underscore the need for improved transparency and detection methods.

Keywords

artificial intelligence; large language models; otolaryngology

INTRODUCTION

In healthcare, large language models (LLMs) such as Chatbot Generative Pre-Trained Transformer (ChatGPT) offer vast opportunities for improving study design, expediting data analysis, and streamlining diagnostic processes and treatment recommendations.^{1–3} Although ChatGPT was initially released to the public in November 2022, the introduction of ChatGPT-4 in March 2023 marked a substantial improvement over earlier models.⁴ This updated version demonstrated notable improvements in accuracy and reliability, including enhanced performance on standardized exams and a reduced tendency to produce fabricated information, also known as “hallucinations.”^{4–6} Moreover, ChatGPT-4 includes a function that allows users to upload files, such as documents, spreadsheets, or images, thereby expanding its abilities to generate text based on inputted data.

Studies of ChatGPT usage in otolaryngology-head and neck surgery (OHNS) have evaluated its ability to answer board questions, respond to patient queries, and make clinical decisions.^{7–9} However, there has been limited investigation into the use of ChatGPT in OHNS publications. The International Committee of Medical Journal Editors (ICMJE) recommends that authors disclose and describe any use of AI-assisted technologies including ChatGPT in assisting research and manuscript preparation.¹⁰ Studies of GPTZero and similar detectors in medical literature more broadly have demonstrated its utility in identifying AI-generated texts with an accuracy up to 80%, including in otolaryngology.^{11–13}

However, even with the use of AI-detectors and expert reviewers, ChatGPT-generated content often goes undetected, especially when the text is assessed at larger scale.^{14,15} Furthermore, AI-detectors limit the possibility of quantifying the amount of text that is LLM-generated. Studies indicate that human reviewers are often unable to reliably distinguish between AI- and human-generated text, with performance only slightly better than random chance.^{16,17} For instance, in a study of 50 peer-reviewed abstracts versus 50

ChatGPT-generated abstracts, blinded human reviewers were only able to correctly identify 68% of generated abstracts as being generated.¹⁸

Recent research on large language models (LLMs) has identified corpus-level patterns that differ from typical human writing, offering an alternative approach for detecting AI-generated text. Liang *et al.* introduced a framework that monitors AI-modified text by analyzing population-level patterns rather than individual instances. Their study evaluated abstracts submitted to leading AI conferences and revealed that GPT models overuse certain words more frequently than human writers. Furthermore, they found that conducting these detection analyses based on specific overutilized adjectives was the most effective approach for identifying AI-generated content.¹⁹ Building on this approach, in this study, we utilized the frequency of the top 100 adjectives disproportionately produced by GPT to quantify and identify trends in AI-generated content in OHNS.

METHODS

Article Selection and Data Extraction

We identified 25 US-based and international OHNS journals and downloaded all publications from these journals between November 2022 and July 2024. This list included the top 10 most cited OHNS journals, as identified by Google Scholar, and 15 additional prominent subspecialty journals.²⁰ We included both fully Open Access journals and Hybrid Access journals which publish both Open Access and Subscription only papers. The author guidelines for all 25 journals were reviewed to determine whether the use of generative AI is formally permitted. Impact factors for each journal were also collected in September 2024. See Table 1 for a list of journals and their publication information.

We then reviewed all publications to exclude Letters to the Editor, Corrections, Comments, Replies, and Consensus Statements. All other publication types, including Original Investigations, Review Articles, Case Reports, Perspectives, etc. were included. Then, using ChatGPT-4, we developed a Python program in Visual Studio Code (Version 1.90.0 (Universal)) to scan the downloaded articles and generate a Microsoft Excel spreadsheet (Microsoft Office 2024, Microsoft, Redmond, United States) reporting the frequency of the 100 adjectives disproportionately used by LLMs in each publication.¹⁹ Then, 5% of articles from each journal were randomly selected and manually verified to ensure the script accurately calculated word counts. Furthermore, articles in which a certain adjective was particularly overexpressed (>10 occurrences) were excluded, as it was determined that this was an indication that the adjective was part of the study's main themes. We manually recorded the country of origin, and the highest degree obtained by the first author for each publication included in the study. After data collection, we excluded articles from countries where ChatGPT is not available, as indicated by the OpenAI website (see Supplementary Table 1).

Statistical Analysis

We used the Mann-Kendall test and logistic regression to assess the trend in the use of GPT-associated adjectives over the study period. The month of the journal issue served

as a proxy for publication date; articles from journal issues spanning multiple months (e.g., “March/April 2024”) were excluded from the time analysis. Mean word count per article pre-March 2023 and post-March 2023 were analyzed using a paired t-test. Logistic regression was employed to examine the association between journal impact factor, as of September 2024, and the frequency of GPT-associated adjective use. For the country analysis, a Mann-Whitney U Test was conducted to compare differences between English-speaking and non-English-speaking countries, while ANOVA analyzed differences among articles from various geographic regions. Analysis of Covariance (ANCOVA) testing was used to adjust for possible confounders. In the author degree analysis, an unpaired t-test was performed to evaluate differences between articles written by first authors with a doctoral degree and those without.

RESULTS

Characteristics of the Included Studies

Applying our primary inclusion and exclusion criteria, we analyzed 25 OHNS journals, identifying 7,134 original articles published between March 2023 and July 2024 (Table 1). After excluding articles from countries where ChatGPT is unavailable, 6,325 articles met the criteria for inclusion in our study. Of these, the majority (66.9%) were published in general ENT journals, defined as those covering two or more subspecialties. The most frequently represented subspecialty journals were Otolaryngology (n = 781; 12.3%), Head and Neck (n = 364; 5.8%), Pediatrics (n = 329; 5.20%), Laryngology (n = 307; 4.85%), Rhinology (n = 184; 2.90%), and Facial Plastic and Reconstructive Surgery (n = 125; 1.98%).

Additionally, using the same 25 journals, we collected 1,617 articles published between November 2022 and February 2023. After excluding articles from countries where ChatGPT is unavailable, 1,377 articles from this period met inclusion criteria. Similarly, a majority were published between general ENT journals (n = 810; 58.8%), followed by Otolaryngology (n = 220; 16.0%), Pediatrics (n = 102; 7.40%), Rhinology (n = 87; 6.32%), Head and Neck (n = 80; 5.81%), Laryngology (n = 61; 4.42%), and Facial Plastic and Reconstructive Surgery (n = 17; 1.23%).

Impact factors, updated as of September 2024, were obtained from the Observatory of International Research (OOIR).²¹ Of the 25 journals, 18 (72%) had documented impact factors and were included in our analysis. We observed a weak but statistically significant negative correlation between impact factor and GPT-associated adjective use (correlation coefficient = -0.15; p < 0.001). In comparisons between fully Open Access journals and Hybrid Access or Subscription-based journals, no significant differences in GPT-associated adjective use were identified.

Time Analysis

Articles that were written prior to March 2023 were included to establish a baseline for adjective usage prior to the public release of ChatGPT-4. When comparing articles pre-March 2023 to post-March 2023, there was significant difference in mean adjective usage per article (5.62 versus 6.01; p = 0.0031).

Articles post-March 2023 were initially categorized into three groups based on their publication dates: the first half of 2023 (March to June), the second half of 2023 (July to December), and the first half of 2024 (January to July). A significant difference was observed in the frequency of GPT-associated adjectives across these three periods (5.56, 5.66, and 6.09 words per article, respectively; $p < 0.001$).

Articles were then sorted by publication date, using journal issue month as a proxy. For this analysis, articles from journal issues covering multiple months (e.g., “March/April 2024”) were excluded. The Mann-Kendall test identified a significant upward trend in GPT-generated adjective use over time ($p < 0.001$). An Ordinary Least Squares (OLS) regression analysis also demonstrated a significant increase in GPT-associated adjective use over time ($p = 0.0141$; Figure 1).

Country and Geographic Region Analyses

Of the 6,325 articles published after March 2023 that met inclusion criteria, 3,526 (55.7%) had first authors from countries where English is the primary language, while 2,799 (44.3%) had first authors from countries where English is not the primary language. Across individual journals and in the overall dataset, authors from English-speaking countries used a significantly higher number of GPT-associated adjectives compared to their counterparts from non-English-speaking countries (median: 4.0 vs. 3.0, $p < 0.001$). This relationship remained significant when controlling for journal impact factor and first author education level ($p < 0.001$). Detailed data on GPT-associated adjective use, stratified by the authors’ country affiliations, is provided in Table 2.

Additionally, articles were categorized based on the geographic region of the country of authorship: Africa, Asia, Europe, Middle East, North America, South America, Oceania, and Multiple. North America was the most represented region with 3,074 articles (48.62%), followed by Europe with 1,458 articles (23.06%) and Asia with 1,244 articles (19.67%). Comprehensive data stratified by geographic region is presented in Table 3. A one-way ANOVA test revealed a statistically significant effect of geographic region on GPT-associated adjective use (F-value: 17.543, $p = 0.000$). To further investigate this effect, a Tukey’s HSD post-hoc test was conducted. Oceania was found to have a significantly higher GPT-associated adjective use compared to both Africa ($p = 0.0087$) and North America ($p = 0.0036$). Conversely, Asia exhibited consistently lower GPT-associated adjective use compared to Europe ($p < 0.001$), the Middle East ($p < 0.001$), North America ($p < 0.001$), and Oceania ($p < 0.001$). When controlling for impact factor, geographic region remained a significant predictor of GPT-associated adjective use overall, but the individual region differences were not significant.

Author Degree Analysis

Of the 6,325 articles published post-March 2023 written in countries where ChatGPT is available, only 3,057 articles (48.3%) included information on the educational degree(s) attained by the first author. Degrees were classified into three levels based on the highest degree attained, doctoral level degree (76.48%), bachelor’s level degree (19.79%), and master’s level degree (3.73%). Articles authored by individuals with doctoral degrees used

slightly fewer GPT-associated adjectives compared to those with non-doctoral degrees (5.02 words per article vs. 5.30), though this difference was not statistically significant ($p = 0.13$). An ANOVA test indicated a statistically significant difference among the degree groups (F-statistic: 6.407, $p = 0.0017$). Post-hoc analysis using Tukey's HSD test identified significant differences between bachelor's and master's level authors (5.07 vs. 6.51, respectively) and between the doctoral and master's level authors (5.02 vs. 6.51, respectively). No significant difference was observed between the bachelor's and doctoral level degree groups.

DISCUSSION

This study aimed to assess trends in the use of AI-generated text in OHNS publications drafting following the public release of ChatGPT-4 in March 2023, using GPT-associated adjective use as a validated model. As described by Liang *et al.*, the selection of words identified in the study was based on an empirical analysis of corpus-level linguistic features of AI-generated and human-written texts.¹⁹ The selection process was then validated using systematic experiments demonstrating that the identified adjectives showed significant frequency spikes in AI-generated reviews after the release of ChatGPT-4. Adjectives were shown to exhibit greater stability than other parts of speech such as adverbs, verbs, and nouns. Moreover, Sun *et al.* (2025) demonstrated that idiosyncrasies between LLMs are primarily attributable to word-level distributions, rather than those in letters or syntax.²² This finding is further corroborated by their predictive model, which exhibits a 97% accuracy in discerning the LLM source of generated text, thus underscoring the validity of employing characteristic word choices to distinguish AI-generated content.²²

Although AI-generated text detectors such as GPTzero have been developed to identify AI-generated content and mitigate the risks associated with its publication, questions remain regarding their accuracy and reliability.¹¹ Notably, Liang *et al.* demonstrated that these detectors show bias against non-native English speakers, especially those with limited vocabulary and linguistic variety. In their study, over half of the non-native English writing samples were misclassified as AI-generated.²³ The authors attribute this issue to the detector's reliance on "text perplexity," which measures how predictable a language model finds the next word in a sentence. Low text perplexity indicates ease in word prediction, while high text perplexity reflects difficulty.

By analyzing the frequency of GPT-associated adjective use, our model effectively avoids the confounding factor of text perplexity. Unlike other studies that use automated detectors to evaluate AI-generated text, we found that authors from English-speaking countries exhibited a higher incidence of AI-generated adjectives than authors from non-English-speaking countries. This disparity may be attributed to the higher frequency of ChatGPT use in the United States, which accounts for the largest global share of traffic to the website (12.78%).²⁴ Future research should focus on assessing the accuracy of AI detection models while addressing potential biases against non-native English speakers.

Our study also observed an increased incidence of GPT-associated adjective use over time, aligning with findings from studies evaluating AI-generated text using automated detectors before and after 2022.¹³ We also identified a small but highly significant difference

between GPT-associated adjective use before and after the public release of ChatGPT-4 in March 2023, demonstrating the sensitivity of this approach in detecting subtle shifts in AI-generated text.

This trend of increasing LLM usage demands awareness of ethical and governance considerations as well as ongoing conversation about how to best navigate the challenges and opportunities associated with these tools. As highlighted in other research, LLMs are susceptible to “chatbot hallucinations,” which lead to the generation of inaccurate information and fictitious references.^{25–27} A study by Májovsky et al. found that GPT-3 can produce a highly convincing fraudulent article that resembles a genuine scientific paper with regard to its structure, use of medical terminology, and overall composition, yet contains specific mistakes, particularly in the references.²⁸ Furthermore, LLM dependency on training data may lead to biases due to errors or outdated information present in the training datasets.^{29,30} Most journals have established detailed guidelines regarding the usage of AI in manuscript preparation and authorship. These guidelines generally state that such tools can be used to improve clarity and readability, as long as authors carefully check the accuracy of the output and recognize their use in the Methods or Acknowledgments section. Moreover, the inclusion of generative AI tools as contributing authors is expressly prohibited. These risks in the setting of globally increased AI usage underscores the need for a thorough evaluation of all articles before publication. Although LLM usage in scientific writing presents certain risks, there are also salient benefits. LLMs can streamline the research process by brainstorming ideas, searching the existing literature, and enhancing the contextualization of findings.³¹ These tools can also expedite data analysis by providing guidance to human researchers and supporting code generation for statistical tests or graphically display results. LLMs also offer valuable assistance to non-native English speakers, helping them articulate their ideas clearly and effectively.³² By breaking down language barriers, LLMs can foster greater inclusivity and collaboration in the global scientific community, enabling contributions from diverse voices across cultures, languages, and educational backgrounds.³²

Limitations

While our study employs a validated model to evaluate AI-generated text at scale, this model only considers adjective usage and does not account for other factors. AI-generated text often has more subtle features such as sentence structure, syntax, punctuation usage, and overall text fluency.

We included articles published from November 2022 through July 2024. However, many articles are conceptualized and written before their print publication. The inability to precisely determine when an article was written represents another limitation of our study.

In analyzing whether the highest degree attained by the first author is significantly associated with GPT-generated text, we categorized author degrees into two groups: doctoral versus non-doctoral degrees. This binary classification simplifies the hierarchy seen particularly in US medicine, where different levels of training, such as a resident physician, fellow, and attending physician, all hold the same M.D. degree.

It should also be noted that this study only focused on ChatGPT-4, which was released in March 2023. OpenAI released ChatGPT-4o in May 2024, which replaced GPT-4 as the default model in July 2024. Our results may not apply to this newer model as we do not have data indicating the most commonly used adjectives with this newer version.

CONCLUSION

This study provides an alternative approach for identifying AI-generated text by focusing on disproportionately used adjectives rather than relying on automated AI detectors, which have been found to be unreliable and biased against non-native English speakers. Our findings reveal that GPT-associated adjective use has increased in frequency over time, has shown an inverse correlation with journal impact factor, and has been more prevalent among authors from English-speaking countries. In a new era of expanding LLMs use and functionality, it is crucial to understand their impact on medical writing and study the quality of the evidence presented in LLM-aided literature.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding:

Anaïs Rameau was supported by a Paul B. Beeson Emerging Leaders Career Development Award in Aging (K76 AG079040) from the National Institute on Aging and by the Bridge2AI award (OT2 OD032720) from the NIH Common Fund.

REFERENCES

1. Shah NH, Entwistle D, Pfeffer MA. Creation and Adoption of Large Language Models in Medicine. *JAMA*. 2023;330(9):866–869. doi:10.1001/jama.2023.14217 [PubMed: 37548965]
2. Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthc J*. 2021;8(2):e188–e194. doi:10.7861/fhj.2021-0095 [PubMed: 34286183]
3. Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine. *Commun Med (Lond)*. 2023;3:141. doi:10.1038/s43856-023-00370-1 [PubMed: 37816837]
4. Open AI, Achiam J, Adler S, et al. GPT-4 Technical Report. Published online March 4, 2024. doi:10.48550/arXiv.2303.08774
5. Kochanek K, Skarzynski H, Jedrzejczak WW. Accuracy and Repeatability of ChatGPT Based on a Set of Multiple-Choice Questions on Objective Tests of Hearing. *Cureus*. 16(5):e59857. doi:10.7759/cureus.59857
6. Massey PA, Montgomery C, Zhang AS. Comparison of ChatGPT–3.5, ChatGPT-4, and Orthopaedic Resident Performance on Orthopaedic Assessment Examinations. *J Am Acad Orthop Surg*. 2023;31(23):1173–1179. doi:10.5435/JAAOS-D-23-00396 [PubMed: 37671415]
7. Hoch CC, Wollenberg B, Lüers JC, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. *Eur Arch Otorhinolaryngol*. 2023;280(9):4271–4278. doi:10.1007/s00405-023-08051-4 [PubMed: 37285018]
8. Zalzal HG, Abraham A, Cheng J, Shah RK. Can ChatGPT help patients answer their otolaryngology questions? *Laryngoscope Investig Otolaryngol*. 2024;9(1):e1193. doi:10.1002/lio2.1193

9. Long C, Lowe K, Zhang J, et al. A Novel Evaluation Model for Assessing ChatGPT on Otolaryngology-Head and Neck Surgery Certification Examinations: Performance Study. *JMIR Med Educ.* 2024;10:e49970. doi:10.2196/49970
10. ICMJE | Recommendations. Accessed September 11, 2024. <https://www.icmje.org/recommendations/>
11. Habibzadeh F GPTZero Performance in Identifying Artificial Intelligence-Generated Medical Texts: A Preliminary Study. *J Korean Med Sci.* 2023;38(38):e319. doi:10.3346/jkms.2023.38.e319 [PubMed: 37750374]
12. Lee PY, Salim H, Abdullah A, Teo CH. Use of ChatGPT in medical research and scientific writing. *Malays Fam Physician.* 2023;18:58. doi:10.51866/cm0006 [PubMed: 37814667]
13. Carnino JM, Chong NYK, Bayly H, Salvati LR, Tiwana HS, Levi JR. AI-generated text in otolaryngology publications: a comparative analysis before and after the release of ChatGPT. *Eur Arch Otorhinolaryngol.* Published online July 17, 2024. doi:10.1007/s00405-024-08834-3
14. Sikander B, Baker JJ, Deveci CD, Lund L, Rosenberg J. ChatGPT-4 and Human Researchers Are Equal in Writing Scientific Introduction Sections: A Blinded, Randomized, Non-inferiority Controlled Study. *Cureus.* 2023;15(11):e49019. doi:10.7759/cureus.49019
15. Pan ET, Florian-Rodriguez M. Human vs machine: identifying ChatGPT-generated abstracts in Gynecology and Urogynecology. *Am J Obstet Gynecol.* 2024;231(2):276.e1–276.e10. doi:10.1016/j.ajog.2024.04.045
16. Else H Abstracts written by ChatGPT fool scientists. *Nature.* 2023;613(7944):423. doi:10.1038/d41586-023-00056-7 [PubMed: 36635510]
17. Clark E, August T, Serrano S, Haduong N, Gururangan S, Smith NA. All That's "Human" Is Not Gold: Evaluating Human Evaluation of Generated Text. Published online July 7, 2021. doi:10.48550/arXiv.2107.00061
18. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digit Med.* 2023;6:75. doi:10.1038/s41746-023-00819-6 [PubMed: 37100871]
19. Liang W, Izzo Z, Zhang Y, et al. Monitoring AI-Modified Content at Scale: A Case Study on the Impact of ChatGPT on AI Conference Peer Reviews. Published online June 15, 2024. doi:10.48550/arXiv.2403.07183
20. Otolaryngology - Google Scholar Metrics. Accessed September 11, 2024. https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=med_otoaryngology
21. Otorhinolaryngology: Journal Rankings | OOIR. Accessed September 12, 2024. <https://oair.org/journals.php?field=Clinical+Medicine&category=Otorhinolaryngology&metric=jif>
22. Sun M, Yin Y, Xu Z, Kolter JZ, Liu Z. Idiosyncrasies in Large Language Models. Published online February 17, 2025. doi:10.48550/arXiv.2502.12150
23. Liang W, Yuksekgonul M, Mao Y, Wu E, Zou J. GPT detectors are biased against non-native English writers. *Patterns (N Y).* 2023;4(7):100779. doi:10.1016/j.patter.2023.100779
24. Singh S ChatGPT Statistics (SEP. 2024) – 200 Million Active Users. DemandSage. September 2, 2024. Accessed September 11, 2024. <https://www.demandsage.com/chatgpt-statistics/>
25. Colasacco CJ, Born HL. A Case of Artificial Intelligence Chatbot Hallucination. *JAMA Otolaryngol Head Neck Surg.* 2024;150(6):457–458. doi:10.1001/jamaoto.2024.0428 [PubMed: 38635259]
26. Alkaissi H, McFarlane SI. Artificial Hallucinations in ChatGPT: Implications in Scientific Writing. *Cureus.* 15(2):e35179. doi:10.7759/cureus.35179
27. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High Rates of Fabricated and Inaccurate References in ChatGPT-Generated Medical Content. *Cureus.* 15(5):e39238. doi:10.7759/cureus.39238
28. Májovský M, Černý M, Kasal M, Komarc M, Netuka D. Artificial Intelligence Can Generate Fraudulent but Authentic-Looking Scientific Medical Articles: Pandora's Box Has Been Opened. *J Med Internet Res.* 2023;25:e46924. doi:10.2196/46924
29. Harrer S Attention is not all you need: the complicated case of ethically using large language models in healthcare and medicine. *eBioMedicine.* 2023;90:104512. doi:10.1016/j.ebiom.2023.104512

30. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. *Lancet Digit Health*. 2023;5(6):e333–e335. doi:10.1016/S2589-7500(23)00083-3 [PubMed: 37120418]
31. Almarie B, Teixeira PEP, Pacheco-Barrios K, Rossetti CA, Fregni F. Editorial – The Use of Large Language Models in Science: Opportunities and Challenges. *Princ Pract Clin Res*. 2023;9(1):1–4. doi:10.21801/ppcrj.2023.91.1 [PubMed: 37693832]
32. Koller D, Beam A, Manrai A, et al. Why We Support and Encourage the Use of Large Language Models in NEJM AI Submissions. *NEJM AI*. 2024;1(1):A1e2300128. doi:10.1056/A1e2300128

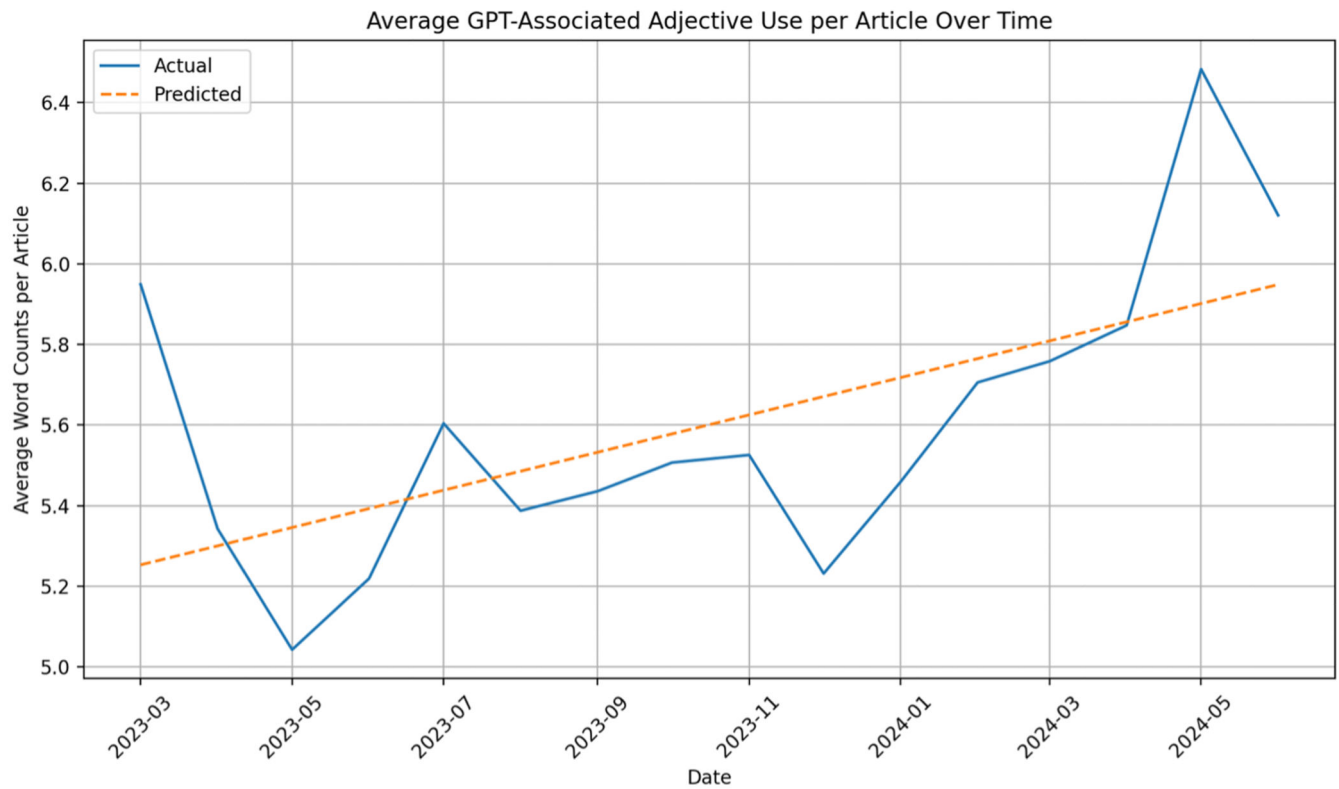


Figure 1. Regression analysis of mean GPT-associated adjective use per article over time. Regression analysis of the mean GPT-associated adjective use per article over time, illustrating a significant upward trend from March 2023 to July 2024. The solid line represents actual adjective usage, while the dashed line represents the predicted trend, highlighting the increasing prevalence of AI-generated text in otolaryngology publications.

Table 1:
List of Journals Used and Respective Journal Information

Journal	OOIR Impact Factor	Official Journal Affiliation(s)	Publication Country	Access Type	First Year of Publication	Published Monthly?	Use of Generative AI Permitted?	# of Publications Included*
American Journal of Otolaryngology	1.8	N/A ***	USA	Hybrid	2018	No	Yes	363
Annals of Otolaryngology, Rhinology & Laryngology	1.3	American Broncho-Esophagological Association	USA	Hybrid	1897	Yes	Yes	262
Auris Nasus Larynx	1.6	Oto-Rhino-Laryngological Society of Japan, Inc.	Netherlands	Hybrid	1973	No	Yes	182
Ear and Hearing	2.6	American Auditory Society	USA	Hybrid	1980	No	Yes	147
Ear, Nose & Throat Journal	1	N/A ***	USA	Open Access	1976	Yes	Yes	311
European Archives of Oto-Rhino-Laryngology	1.9	European Federation of Oto-Rhino-Laryngological Societies	Germany	Hybrid	1990	Yes	Yes	665
Facial Plastic Surgery & Aesthetic Medicine	N/A **	American Academy of Facial Plastic and Reconstructive Surgery	USA	Hybrid	2020	No	Yes	125
Frontiers Otorhinolaryngology-Head and Neck Surgery	N/A **	N/A ***	Switzerland	Open Access	2015	No	Yes	30
Head & Neck	N/A **	International Academy of Oral Oncology	USA	Hybrid	1978	Yes	Yes	364
Hearing Research	2.5	N/A ***	Sweden	Hybrid	1979	Yes	Yes	146
International Forum of Allergy & Rhinology	6.4	American Rhinologic Society American Academy of Otolaryngic Allergy	USA	Hybrid	2011	Yes	Yes	184
International Journal of Otorhinolaryngology and Head and Neck Surgery	N/A **	N/A ***	India	Open Access	2017	No		197
International Journal of Pediatric Otorhinolaryngology	1.2	N/A ***	Ireland	Hybrid	1979	Yes	Yes	329
JAMA Otolaryngology-Head & Neck Surgery	6	American Head and Neck Society	USA	Hybrid	2013	Yes	Yes	223
Journal of Voice	2.5	Voice Foundation, The International	USA	Hybrid	1987	No	Yes	307

Journal	OOIR Impact Factor	Official Journal Affiliation(s)	Publication Country	Access Type	First Year of Publication	Published Monthly?	Use of Generative AI Permitted?	# of Publications Included*
		Association of Phonosurgery						
Laryngoscope Investigative Otolaryngology	1.6	The Triological Society	USA	Open Access	2015	No	Yes	205
MDPI - Journal of Otorhinolaryngology, Hearing and Balance Medicine	N/A **	N/A ***	Switzerland	Open Access	2018	No	Yes	16
ORL (Karger)	1.5	N/A ***	Switzerland	Hybrid	1938	No	Yes	46
Otolaryngology–Head and Neck Surgery	2.6	American Academy of Otolaryngology–Head and Neck Surgery Foundation	USA	Hybrid	1981	Yes	Yes	513
OTO Open	N/A **	American Academy of Otolaryngology–Head and Neck Surgery Foundation	USA	Open Access	2017	No	Yes	100
Otology & Neurotology	1.9	American Otological Society American Neurotology Society	USA	Hybrid	2001	No	Yes	373
Otology & Neurotology Open	N/A **	American Otological Society American Neurotology Society	USA	Open Access	2021	No	Yes	20
The Journal of International Advanced Otology	1	European Academy of Otology and Neurotology The Politzer Society	Turkey	Open Access	2009	No	Yes	95
The Journal of Laryngology & Otology	1.1	N/A ***	United Kingdom	Hybrid	1887	Yes	N/A	286
The Laryngoscope	2.2	The Triological Society	USA	Hybrid	1896	Yes	Yes	836

* Only reflects publications from countries in which ChatGPT is available

** No Impact Factor available listed on the Observatory of International Research (OOIR)

*** No societal or organizational affiliation

Table 2:
Analysis of GPT-Associated Adjective Usage by Authors’ Country Affiliations (March 2023 – July 2024)

Journal	Number of Papers (%)		Mean Word Count		p-Value
	English	Non-English	English	Non-English	
American Journal of Otolaryngology	233 (64.2%)	130 (35.8%)	6.73	5.85	0.073
Annals of Otolology, Rhinology & Laryngology	210 (80.2%)	52 (19.8%)	6.24	5.00	0.019 *
Auris Nasus Larynx	9 (4.9%)	173 (95.1%)	7.11	4.79	0.148
Ear and Hearing	103 (70.1%)	44 (29.9%)	10.86	11.80	0.483
Ear, Nose & Throat Journal	115 (37.0%)	196 (63.0%)	3.57	2.84	0.053
European Archives of Oto-Rhino-Laryngology and Head and Neck	56 (8.4%)	608 (91.6%)	6.71	6.50	0.784
Facial Plastic Surgery & Aesthetic Medicine	102 (81.6%)	23 (18.4%)	3.89	4.17	0.756
Frontiers Otorhinolaryngology - Head and Neck Surgery	3 (10.0%)	27 (90.0%)	7.67	9.26	0.728
Head and Neck	183 (50.3%)	181 (49.7%)	6.99	6.78	0.662
Hearing Research	80 (54.8%)	66 (45.2%)	9.75	8.08	0.184
International Forum of Allergy and Rhinology	155 (84.2%)	29 (15.8%)	2.70	1.97	0.143
International Journal of Otolaryngology	5 (2.5%)	192 (97.5%)	5.20	3.90	0.472
International Journal of Pediatric Otorhinolaryngology	211 (64.1%)	118 (35.9%)	7.08	7.43	0.534
JAMA Otolaryngology – Head & Neck Surgery	194 (87.4%)	28 (12.6%)	3.43	3.07	0.532
Journal of Voice	119 (38.9%)	187 (61.1%)	6.19	5.43	0.154
Laryngoscope Investigative	119 (58.0%)	86 (42.0%)	6.87	6.10	0.284
MDPI Otolaryngology	3 (18.75%)	13 (81.25%)	11.33	7.00	0.231
ORL - Karger Publishers	16 (34.8%)	30 (65.2%)	7.56	3.63	0.026 *
Otolaryngology-Head and Neck Surgery	409 (79.7%)	104 (20.3%)	6.35	6.66	0.552
Otolaryngology-Head and Neck Surgery Open	84 (84.0%)	16 (16.0%)	5.85	5.75	0.935
Otology & Neurotology	241 (64.6%)	132 (35.4%)	5.77	6.07	0.556
Otology & Neurotology Open	16 (80.0%)	4 (20.0%)	5.75	2.25	0.015 *
The Journal of International Advanced Otology	9 (9.5%)	86 (90.5%)	7.11	4.05	0.122
The Journal of Laryngology and Otology	178 (62.2%)	108 (37.8%)	5.40	4.50	0.038 *
The Laryngoscope	671 (80.3%)	165 (19.7%)	5.87	4.27	< 0.001 *

* = significant value based on $\alpha=0.95$
Only includes papers that were written by authors in countries where ChatGPT-4 is available

Table 3:
Comprehensive Analysis of Data Stratified by Geographic Region

Geographic Region	Mean Number of GPT-Associated Adjectives	Number of Publications from the region (%)
Africa	1.806	31 (0.49)
Asia	2.333	1244 (19.67)
Europe	3.94	1458 (23.06)
Middle East	4.037	218 (3.45)
Multiple	4.848	46 (0.73)
North America	3.579	3074 (48.62)
Oceania	5.212	132 (2.09)
South America	3.653	120 (1.9)