

ORIGINAL RESEARCH ARTICLE

Neck dissection in head and neck surgery: An assessment of ChatGPT performance

Dustin A. Silverman^{1*}, John S. Howard¹, Priscilla F. A. Pichardo¹,
Yash J. Patil¹, Mekibib Altaye², Chad A. Zender¹, and Alice L. Tang¹¹Department of Otolaryngology, Head and Neck Surgery, University of Cincinnati, Cincinnati, Ohio, United States of America²Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center, University of Cincinnati College of Medicine, Cincinnati, Ohio, United States of America

Abstract

Artificial intelligence models such as chat generative pre-trained transformer (ChatGPT) are being increasingly used to inform treatment-related decisions. Among otolaryngology subspecialties, there is a paucity of literature examining the role of ChatGPT within head and neck surgical oncology. The utility of ChatGPT in addressing questions related to surgically relevant anatomy and lymphadenectomy procedures remains poorly understood. The primary pilot study objective was to determine the reliability of ChatGPT in answering neck dissection-related inquiries compared to expert head and neck surgical oncologists. Five neck dissection-related questions were presented to ChatGPT v3.5. Three fellowship-trained head and neck surgeons compared AI-generated responses to those of an expert head and neck surgeon. Raters, blinded to the author's identity, evaluated the responses given based on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The median level of agreement between raters for the ChatGPT responses was 1.0 (interquartile range [IQR]: 1.0, 2.5; minimum = 1 and maximum = 4), while the median level of agreement between raters for the surgeon responses was 5.0 (IQR: 5.0, 5.0; minimum = 5 and maximum = 5). The Mann-Whitney *U* test yielded a significance level of $p=0.007$ when comparing the level of agreement between ChatGPT and surgeon responses. Raters showed minimal consistency when evaluating ChatGPT responses (intraclass correlation coefficient = 0.05; 95% confidence interval: 0.0–0.88), in contrast to perfect agreement observed for the surgeon responses. In summary, ChatGPT is a promising tool in the acquisition of surgical knowledge. For neck dissection-related inquiries, a discrepancy between the reliability of ChatGPT-generated responses and surgeon expertise exists. Further refinement in AI models is needed to strengthen the utility of ChatGPT in head and neck oncologic surgery.

***Corresponding author:**Dustin A. Silverman
(silverdt@ucmail.uc.edu)

Citation: Silverman DA, Howard JS, Pichardo PFA, *et al.* Neck dissection in head and neck surgery: An assessment of ChatGPT performance. *Artif Intell Health.* 2026;3(2):025240053. doi: 10.36922/AIH025240053

Received: June 9, 2025**Revised:** September 1, 2025**Accepted:** September 11, 2025**Published online:** November 25, 2025**Copyright:** © 2025 Author(s).

This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Keywords: Artificial intelligence; Chat generative pre-trained transformer; Head and neck; Neck dissection; Lymphadenectomy

1. Introduction

The interest in applications of artificial intelligence (AI) in medicine and surgery has increased exponentially since the emergence of chat generative pre-trained transformer (ChatGPT) developed by OpenAI in 2022. ChatGPT is a large language

model (LLM) trained on vast amounts of data, including medical literature, which allows it to provide answers to queries from an exceedingly diverse variety of sources.¹ As one of the largest publicly available LLMs, ChatGPT has established itself as the preeminent AI-based resource among physicians, surgeons, and scientists, with an estimated and growing user base of 100 million.²

There has been an increased focus and attention to ChatGPT applications in otolaryngology, given its potential in medical knowledge acquisition, research question formulation, and medical education improvement, which holds promise in advancing otolaryngology practice more broadly.³ Based on a state-of-the-art review by Lechien *et al.*,⁴ the AI model is more accurate in providing diagnoses rather than suggesting the most adequate additional examinations and treatments when presented with real clinical cases. Further, ChatGPT's primary applications within the field are currently limited to generating disease or treatment information, which are subject to wide variability and potential error. The integrity and fidelity of AI tools in healthcare depend on their ability to provide accurate and well-informed responses. Thus, inaccurate outputs, especially in the context of surgery, may undermine patient and provider trust as a supplementary resource. Despite this, ChatGPT and similar LLM programs carry great promise as an adjunctive clinical tool for improving clinical otolaryngology practice.⁵

To date, there has been a paucity of investigations focusing on ChatGPT's application to the subspecialty of head and neck surgical oncology, with most focusing on preoperative counseling and alignment with the National Comprehensive Cancer Network (NCCN) guideline-recommended therapy.⁶⁻¹⁰ Maniyar *et al.*¹¹ investigated ChatGPT's ability to evaluate a dataset of 25 oral cavity malignancy questions and demonstrated an overall accuracy rate of 80% in providing comprehensive and correct responses for 20 questions; while the highest accuracy was recorded for questions related to basic oral cavity cancer and prevention, those queries relating to specific treatment and post-operative aspects prompted a mix of partially correct and incomplete responses. Although these studies are effective in establishing the possibilities and limitations of the emerging technology, such findings highlight the current challenges and pitfalls associated with its use.

Importantly, there is a lack of understanding regarding ChatGPT's knowledge as it relates to lymphadenectomy procedures, anatomy, and nuances in guideline-directed treatment recommendations. To address this gap in knowledge, our primary study objective was to determine the reliability of ChatGPT in answering neck

dissection-related inquiries compared to expert head and neck surgical oncologists.

2. Methods

A pilot study was conducted to investigate the level of agreement in responses given by an expert, fellowship-trained head and neck surgical oncologist, and ChatGPT to common lymphadenectomy (neck dissection) procedural questions and to examine the ability of ChatGPT in identifying relevant surgical anatomy. The current study was deemed exempt from Institutional Review Board approval given the absence of patient-level data needed to conduct the research. The authors have no disclosures or affiliations related to ChatGPT or its developer, OpenAI Inc. A total of three board-certified, fellowship-trained head and neck surgical oncologists and reconstructive surgeons were selected to answer questions based on their amount of clinical expertise in the field.

ChatGPT version 3.5 was selected as the AI platform for this pilot study, which was introduced in November 2022. This version, rather than the later iterations which require a fee-based subscription, was selected as it is freely accessible to users. Given this reason, the financial barriers to use are eliminated and allow users of various backgrounds and financial situations to interact with the platform, ensuring a wider range of operator access, diversity, and inclusivity.

A total of five neck dissection-related questions were developed and presented to ChatGPT version 3.5 on February 25, 2024. To ensure consistency, all questions were entered independently into the ChatGPT engine using the "New Chat" function by a single investigator. The same questions were simultaneously answered by a practicing, academic, fellowship-trained head and neck surgical oncologist and reconstructive surgeon (DAS). The questions were designed to focus on relevant anatomical considerations and established national treatment guidelines^{12,13} and are outlined as follows:

- (1) What are the anatomic boundaries of a level IA selective neck dissection (END)?
- (2) What anatomic structure separates level IIA from level IIB in a selective neck dissection?
- (3) What anatomic structures are sacrificed in a radical neck dissection?
- (4) What END levels are recommended to include for a patient with cT2N0M0 right lateral tongue squamous cell carcinoma?
- (5) When is an END indicated for a cT1N0M0 right lateral tongue squamous cell carcinoma based on depth of invasion (DOI)?

Following this, the five questions and respective answers were tabulated and distributed to a total of three raters (YJP,

CAZ, and ALT). Notably, the raters were similarly selected based on their significant expertise in the field of head and neck surgical oncology and reconstruction with a combined 49 years of clinical experience (range: 8–23 years). Raters were blinded to the author of the respective responses (e.g., ChatGPT vs. expert fellowship-trained surgeon). Raters graded the accuracy and completeness of the responses for each question using a Likert scale method with values ranging from 1 to 5 (1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, 5 = strongly agree). Grading results are displayed in Table 1, and the study design is displayed in Figure 1.

2.1. Statistical analysis

Quantitative ratings separately obtained from the three raters for the ChatGPT and expert surgeon responses were reported as median and interquartile range (IQR). The difference in mean ratings was evaluated using the Mann–Whitney *U* test. The agreement between raters was evaluated using the intraclass correlation coefficient (ICC). The significance threshold used was $p < 0.05$. All statistical analyses were completed using SAS version 9.0 (SAS Institute, Inc., Cary, NC, USA) statistical software.

3. Results

A total of five neck dissection-specific questions were presented to ChatGPT and an expert head and neck surgeon. The respective responses were subsequently evaluated by raters blinded to the author of each response (Table A1). The median level of agreement between raters for the ChatGPT responses was 1.0 (IQR: 1.0, 2.5; minimum = 1 and maximum = 4), while the median level of agreement between raters for the surgeon responses was 5.0 (IQR: 5.0, 5.0; minimum = 5 and maximum = 5). The Mann–Whitney *U* test yielded a significance level of $p = 0.007$ when comparing the level of agreement between ChatGPT and expert surgeon responses. For the ChatGPT responses, questions 1 and 2 received the

worst ratings (1, strongly disagree) from all the expert raters while questions 3, 4, and 5 received mixed ratings; however, the average scores were low (< 2.6). Raters showed minimal consistency when evaluating ChatGPT responses (ICC = 0.05; 95% confidence interval: 0.0–0.88), in contrast to perfect agreement (5, strongly agree) observed for the expert surgeon responses from all blinded expert raters.

4. Discussion

The introduction of ChatGPT into the medical and surgical landscape has significantly altered how patients and healthcare practitioners interact with and utilize AI to enhance education and treatment decision-making within the field of otolaryngology. To date, there have been a limited number of studies focusing on the utility of ChatGPT in head and neck oncology and reconstructive surgery (Table 2).^{8–11,14} A comprehensive review by Lechien *et al.*⁴ aimed to review the current literature on the application, accuracy, and performance of ChatGPT in otolaryngology at the 1-year anniversary of the platform's public release; fewer than 10 of the 40 included studies focused on the subspecialty of head and neck oncology. Importantly, even fewer analyses of ChatGPT have specifically assessed its knowledge surrounding lymphadenectomy procedures—a cornerstone of head and neck surgery—and surgically relevant anatomy. Thus, its utility in understanding the application and limits of neck dissection procedures remains poorly understood. To the authors' knowledge, this is the first pilot study to explicitly evaluate ChatGPT's reliability in answering neck dissection-related inquiries compared to expert head and neck surgical oncologists.

The current study demonstrated a higher level of agreement regarding the accuracy and completeness of expert surgeon responses compared to ChatGPT in assessing neck dissection-related anatomy and nuanced treatment recommendations ($p = 0.007$). Notably, the mean level of agreement over five surgeon responses between

Table 1. Expert head and neck surgeon Likert scale grading of ChatGPT versus surgeon response

Question	Surgeon 1		Surgeon 2		Surgeon 3	
	ChatGPT response	Surgeon response	ChatGPT response	Surgeon response	ChatGPT response	Surgeon response
Question 1	1	5	1	5	1	5
Question 2	1	5	1	5	1	5
Question 3	3	5	4	5	1	5
Question 4	1	5	1	5	4	5
Question 5	2	5	1	5	4	5
Total score	8	25	8	25	11	25
Mean score	1.6	5	1.6	5	2.2	5

Notes: 1=strongly disagree, 2=disagree, 3=neither agree nor disagree, 4=agree, 5=strongly agree.

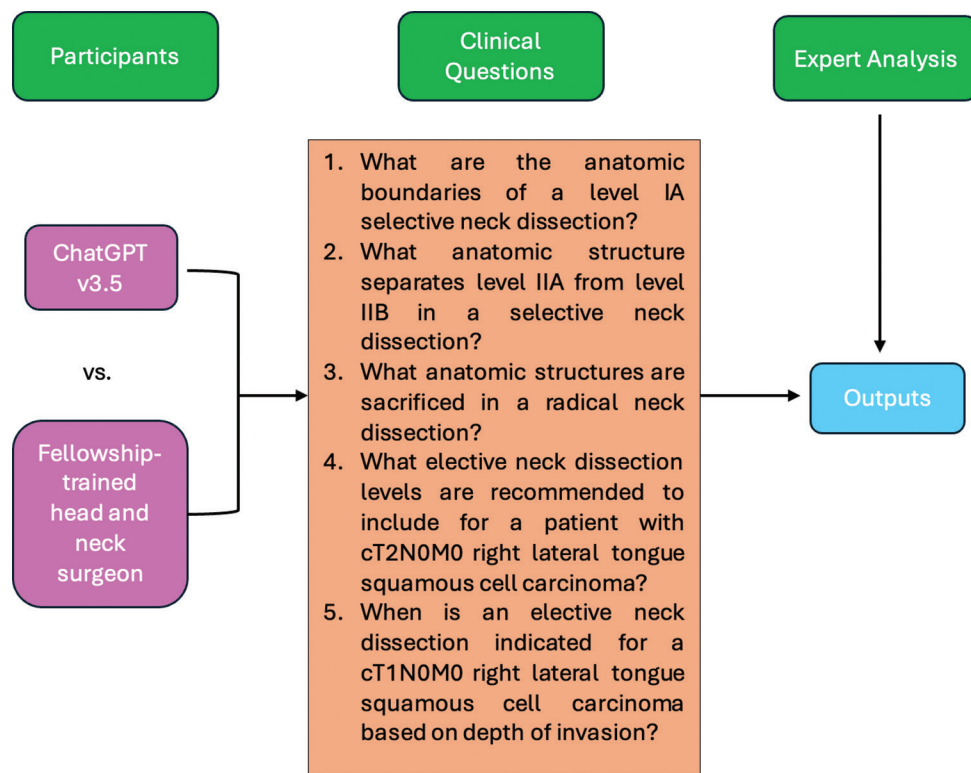


Figure 1. Study design comparing responses from ChatGPT (v3.5) and fellowship-trained head and neck surgeons to a series of clinical questions related to neck dissection. Responses were evaluated through expert analysis to generate comparative outputs.

Table 2. Head and neck-focused ChatGPT studies

Author	Year	ChatGPT version	Reported accuracy and/or rating of head and neck-focused responses	Key findings
Kuscu <i>et al.</i> ⁷	2024	4.0	<ul style="list-style-type: none"> 86.4% rated as “comprehensive/correct” 11% rated as “incomplete/partially correct” 2.6% rated as “mixed data/misleading” 	<ul style="list-style-type: none"> Most questions ranked as “comprehensive/correct” were related to basic knowledge, diagnosis, and/or treatment
Lee <i>et al.</i> ⁸	2023	3.5 and 4.0	<ul style="list-style-type: none"> Chatbot responses were similar in terms of readability, inclusion of medical errors, and knowledge content compared to online patient education materials 	<ul style="list-style-type: none"> Surgeons preferred chatbot responses over online patient education materials 48% of the time
Washington <i>et al.</i> ⁹	2024	3.5	<ul style="list-style-type: none"> 95% of responses were formulated in compliance with the 2021 NCCN treatment guidelines 55% of responses included incorrect staging 	<ul style="list-style-type: none"> 40% of responses included unnecessary treatment recommendations 60% of responses erroneously omitted neck dissection in surgical recommendations
Mnajjed and Patel ¹⁰	2024	Not specified	<ul style="list-style-type: none"> Chatbot responses contained sufficient detail regarding reasons for surgery, steps of surgery, risks, and basic postoperative care 	<ul style="list-style-type: none"> All chatbot responses were ranked at a college-level readability score, which is higher than recommended 8 of 10 surgical complications were not mentioned by the chatbot without further questioning or information
Maniyar <i>et al.</i> ¹¹	2024	3.5	<ul style="list-style-type: none"> 80% of chatbot responses for 20 oral cancer-focused questions were rated as “comprehensive/correct” in terms of accuracy 	<ul style="list-style-type: none"> Question topics with the highest ranked accuracy were basic oral cancer knowledge as well as cancer prevention
Tessler <i>et al.</i> ¹⁴	2024	3.5	<ul style="list-style-type: none"> 59.7% of responses were rated as “highly accurate” in regard to following the recommended AAO-HNS guidelines 	<ul style="list-style-type: none"> Response accuracy varied significantly by otolaryngology subspecialty Head and neck response accuracy was rated as 100%, rhinology and neurotology/otology at 66%, laryngology at 50%, and pediatrics at 8%

Abbreviations: AAO-HNS: American Academy of Otolaryngology–Head and Neck Surgery; ChatGPT: Chat generative pre-trained transformer; NCCN: National comprehensive cancer network.

blinded, expert raters achieved a unanimous decision of “strongly agree” on the Likert scale. Interestingly, questions 1 and 2 which were posed to ChatGPT in an open-ended fashion received the lowest ratings with all raters noting “strong disagreement” with the AI platform’s responses. This is likely due to the universally accepted and clearly defined anatomical boundaries of lymphadenectomy levels among head and neck surgeons, which were incorrectly identified by ChatGPT. For example, ChatGPT was unable to correctly identify the muscular boundaries of a level IA neck dissection. When ChatGPT was further challenged to identify the anatomic structure separating level IIA from level IIB in a selective neck dissection, it correctly recognized the spinal accessory nerve (cranial nerve XI); however, it provided additional, unprompted detail in its response, and alarmingly answered that level IIB “includes the removal of lymph nodes along with the spinal accessory nerve.” Despite the advantages of LLMs as an adjunctive tool in otolaryngology, such responses highlight the pitfalls of the current technology and that broader incorporation in clinical practice should be interpreted with caution. As an LLM is limited by the data on which it was trained, future iterations and refinements may continue to increase the accuracy of these models with the incorporation of specialty-specific platforms designed to optimize accuracy and limit errors, which may jeopardize the safety of patients and negatively impact the decision-making by clinicians.¹⁵

Studies that assess the application of ChatGPT in head and neck surgical oncology have reached similar conclusions. In a study of 40 clinical vignettes designed to identify ChatGPT’s capacity to provide appropriate staging and treatment recommendations for head and neck mucosal malignancies, Washington *et al.* found that while ChatGPT’s responses were 95% accurate in first-line treatment recommendations based on the 2021 NCCN guidelines, 55% contained inaccurate staging.⁹ Similar to the current study, 40% of treatment recommendations were deemed unnecessary, and neck dissection was incorrectly omitted from treatment recommendations in 50% of cases. Moreover, ChatGPT offered extraneous therapeutic options in this study and suggested that unnecessary additional neck levels should be included in the typical treatment of early-stage oral cavity squamous cell carcinoma. Maniyyar *et al.*¹¹ highlighted similar limitations in ChatGPT’s oral cancer knowledge with 40% of the treatment responses categorized as “incomplete/partially correct” with as high as 20% of the responses pertaining to diagnosis judged to be “completely inaccurate/irrelevant.” In contrast to these findings, no “completely inaccurate/irrelevant” responses were observed in 154 head and neck cancer-related questions posed by Kuşcu *et al.*,⁷ though a high proportion of misleading and partially correct responses were still

generated. This underscores ChatGPT’s frequent deviation from established oncologic guidelines, which may lead to exaggerated treatment plans and inaccurate estimations of morbidity for patients and families seeking treatment-related information.

ChatGPT continued to face additional limitations when challenged with more nuanced clinical scenarios and was unable to correctly identify the indications to perform an END for cT1N0M0 right lateral tongue squamous cell carcinoma based on DOI. Specifically, ChatGPT answered that an END should be performed if DOI exceeds 4 mm, which is contrary to high-quality evidence provided by D’Cruz *et al.*,¹⁶ which suggested in a *post hoc* analysis that END resulted in improved oncologic outcomes when tumor DOI was >3 mm. Moreover, NCCN guidelines provide additional, more nuanced recommendations and state that careful clinical judgement should be exercised in cases with DOI of 2–4 mm and that various clinicopathologic factors (e.g., degree of clinical suspicion, patient age, reliability of follow-up for surveillance) should be used in decision-making. Similarly, when presented with a variety of clinical otolaryngology questions based on the guidelines of the American Academy of Otolaryngology, Tessler *et al.*¹⁴ also showed that ChatGPT generated mixed results as the model performed better in certain subspecialties, such as head and neck, while less well in laryngology and pediatrics (100% in head and neck vs. 8.3% in pediatrics). The authors postulate that this may be in part due to the small number of questions generated in some subspecialties and the limited availability of relevant data in the model’s training dataset, which may also explain the decreased accuracy of the responses in this study. Despite its potential drawbacks, ChatGPT may be further used to generate pre-surgical educational information regarding the indications, risks, and recovery of neck dissection procedures and has been shown to be comparable to publicly available websites with roughly equal preference among experienced head and neck surgeons.⁸ As such, AI chatbots may help to address critical barriers that exist in head and neck cancer patient counseling and education for common oncologic surgeries.

Several limitations are noted in this study. First, the narrow question set may limit overall generalizability across the head and neck discipline and the ability of ChatGPT to fully assess lymphadenectomy procedures. Similarly, other LLMs may differ in their responses when challenged with similar questions, though these comparisons are outside the scope of this pilot study. Future work may be designed to assess these differences. Only head and neck surgeons from a single academic institution were included, thus introducing the possibility of institutional bias. As a freely accessible platform, ChatGPT version 3.5 was selected

for use in this study; however, the fee-based subscription version may potentially offer an improvement in responses. Further study is required to clarify these differences. In addition, as with all studies involving chatbots, the risk of “hallucination” events—a phenomenon where inaccurate or nonsensical content is generated—remains possible, and has been demonstrated to occur more commonly in ChatGPT version 3.5 compared to version 4.0.¹⁷ Although reports of these instances in medical and otolaryngology literature are rare, in an evaluation of 75 responses generated by nine different LLMs, De Wynter *et al.*¹⁸ found that 15.4% of the responses contained incorrect claims and/or misattributed published literature. These hallucinations may also include the fabrication of scientific references. In a separate study by Bhattacharyya *et al.*,¹⁹ 47% of 115 references provided by ChatGPT were found to be fabricated. Finally, AI-based platforms and LLMs are limited by the quality and accuracy of input data. Thus, it is important to recognize that the model’s training data may not be fully up-to-date with the most recent treatment guidelines. To overcome this barrier, future iterations may incorporate feedback and input from practicing otolaryngologists to ensure consistency and improve overall performance. Despite these limitations, a primary strength of the current study is its simplicity in design and reproducibility. To the authors’ knowledge, this is the first study to specifically assess ChatGPT’s ability to answer lymphadenectomy-related queries, including procedure-specific knowledge, anatomical boundaries, and guideline-based recommendations in clinical neck dissection scenarios.

5. Conclusion

In conclusion, experienced head and neck surgeon responses outperformed ChatGPT answers in the focused assessment of neck dissection-related queries ($p=0.007$). Although the popular AI platform is a valuable adjunctive source of information for patients and providers, it remains limited in its ability to accurately generate responses to more nuanced head and neck prompts and outputs should continue to be interpreted with caution. Further otolaryngology and guideline-based training are required to refine ChatGPT responses in future iterations; additional study is needed to more clearly define its boundaries in addressing neck dissection-specific knowledge.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Dustin A. Silverman, John S. Howard, Alice L. Tang

Investigation: Dustin A. Silverman, John S. Howard, Alice L. Tang

Methodology: Dustin A. Silverman, John S. Howard, Mekibib Altaye, Alice L. Tang

Writing—original draft: Dustin A. Silverman, John S. Howard, Mekibib Altaye, Alice L. Tang

Writing—review & editing: All authors

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. 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.
doi: 10.3389/frai.2023.1169595
2. Mesko B. The ChatGPT (generative artificial intelligence) revolution has made artificial intelligence approachable for medical professionals. *J Med Internet Res.* 2023;25:e48392.
doi: 10.2196/48392
3. Schmalbach CE. Our otolaryngology future with artificial intelligence. *Otolaryngol Head Neck Surg.* 2024;170(6):1483.
doi: 10.1002/ohn.802
4. Lechien JR, Rameau A. Applications of ChatGPT in otolaryngology-head neck surgery: A state of the art review. *Otolaryngol Head Neck Surg.* 2024;171(3):667-677.
doi: 10.1002/ohn.807
5. Chiesa-Estomba CM, Speth MM, Mayo-Yanez M, Liu DT, Maniaci A, Borsetto D. Is the evolving role of artificial intelligence and chatbots in the field of otolaryngology embracing the future? *Eur Arch Otorhinolaryngol.* 2024;281(4):2179-2180.
doi: 10.1007/s00405-023-08382-2
6. Davis RJ, Ayo-Ajibola O, Lin ME, Swanson MS, Chambers TN, Kwon DI, *et al.* Evaluation of oropharyngeal

- cancer information from revolutionary artificial intelligence chatbot. *Laryngoscope*. 2024;134(5):2252-2257.
doi: 10.1002/lary.31191
7. Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? *Front Oncol*. 2023;13:1256459.
doi: 10.3389/fonc.2023.1256459
8. Lee JC, Hamill CS, Shnayder Y, Buczek E, Kakarala K, Bur AM. Exploring the role of artificial intelligence chatbots in preoperative counseling for head and neck cancer surgery. *Laryngoscope*. 2024;134(6):2757-2761.
doi: 10.1002/lary.31243
9. Washington CJ, Abouyared M, Karanth S, *et al*. The use of chatbots in head and neck mucosal malignancy treatment recommendations. *Otolaryngol Head Neck Surg*. 2024;171:1062-1068.
doi: 10.1002/ohn.818
10. Mnajjed L, Patel RJ. Assessment of ChatGPT generated educational material for head and neck surgery counseling. *Am J Otolaryngol*. 2024;45(5):104410.
doi: 10.1016/j.amjoto.2024.104410
11. Maniyar N, Sarode GS, Sarode SC, Thakkar S. ChatGPT conversations on oral cancer: Unveiling ChatGPT's potential and pitfalls. *Oral Oncol Rep*. 2024;10:100280.
doi: 10.1016/j.oor.2024.100280
12. National Comprehensive Cancer Network. *Head and Neck Cancers*. Ver. 1. Pennsylvania: National Comprehensive Cancer Network; 2024.
13. Miller MC, Goldenberg D. AHNS Series: Do you know your guidelines? Principles of surgery for head and neck cancer: A review of the national comprehensive cancer network guidelines. *Head Neck*. 2017;39(4):791-796.
doi: 10.1002/hed.24654
14. Tessler I, Wolfowitz A, Alon EE, *et al*. ChatGPT's adherence to otolaryngology clinical practice guidelines. *Eur Arch Otorhinolaryngol*. 2024;281(7):3829-3834.
doi: 10.1007/s00405-024-08634-9
15. Long C, Subburam D, Lowe K, *et al*. ChatENT: Augmented Large language model for expert knowledge retrieval in otolaryngology-head and neck surgery. *Otolaryngol Head Neck Surg*. 2024;171(4):1042-1051.
doi: 10.1002/ohn.864
16. D'Cruz AK, Vaish R, Kapre N, *et al*. Elective versus therapeutic neck dissection in node-negative oral cancer. *N Engl J Med*. 2015;373(6):521-529.
doi: 10.1056/nejmoa1506007
17. Chelli M, Descamps J, Lavoué V, *et al*. Hallucination rates and reference accuracy of chatgpt and bard for systematic reviews: Comparative analysis. *J Med Internet Res*. 2024;26:e53164.
doi: 10.2196/53164
18. De Wynter A, Wang X, Sokolov A, Gu Q, Chen SQ. An evaluation on large language model outputs: Discourse and memorization. *Nat Lang Process J*. 2023;4:100024.
19. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus*. 2023;15(5):e39238.
doi: 10.7759/cureus.39238

Appendix

Table A1. Supplementary information

Question	ChatGPT's answer	Surgeon's answer
1. What are the anatomic boundaries of a level IA selective neck dissection?	<p>In a selective neck dissection, Level IA typically involves the removal of lymph nodes in the submental and submandibular triangle. The anatomical boundaries of Level IA selective neck dissection are as follows:</p> <ol style="list-style-type: none"> 1. Superiorly: The inferior border of the mandible 2. Inferiorly: The hyoid bone 3. Anteriorly: The anterior belly of the digastric muscle 4. Posteriorly: The posterior belly of the digastric muscle and the stylomandibular ligament. <p>Level IA selective neck dissection aims to remove lymph nodes in the submental and submandibular regions while preserving other critical structures such as major nerves and blood vessels.</p>	<p>Level IA refers to the lymph nodes within the submental triangle.</p> <p>The anatomic boundaries of a level IA selective neck dissection are:</p> <ol style="list-style-type: none"> 1. Superior: Inferior border of the mandible 2. Inferior: Hyoid bone 3. Lateral: Anterior belly of the digastric muscle 4. Anterior: Platysma muscle 5. Posterior: Mylohyoid muscle
2. What anatomic structure separates level IIA from level IIB in a selective neck dissection?	<p>In a selective neck dissection, the anatomic structure that separates level IIA from level IIB is the spinal accessory nerve (CN XI). This nerve runs through the neck and serves the trapezius and sternocleidomastoid muscles.</p> <p>Level IIA involves the removal of lymph nodes but spares the spinal accessory nerve, while level IIB includes the removal of lymph nodes along with the spinal accessory nerve.</p>	<p>The spinal accessory nerve (CN XI) separates level IIA from level IIB in a selective neck dissection.</p>
3. What anatomic structures are sacrificed in a radical neck dissection?	<p>In a radical neck dissection, the following anatomic structures are typically sacrificed:</p> <ol style="list-style-type: none"> 1. Lymph nodes: Lymph nodes in levels I–V of the neck are removed 2. Sternocleidomastoid muscle: The entire sternocleidomastoid muscle is removed 3. Internal jugular vein: The internal jugular vein is excised or ligated 4. Accessory nerve (spinal accessory nerve, CN XI): The spinal accessory nerve is sacrificed 5. Submandibular gland: The submandibular gland may be removed depending on the extent of the surgery 6. Fat and connective tissue: These tissues are often excised to ensure complete removal of lymph nodes and to prevent recurrence of cancer. <p>Radical neck dissection is a surgical procedure used to treat advanced or metastatic head and neck cancers. It aims to remove all potentially cancerous tissue in the neck region, which may involve sacrificing several important anatomical structures to achieve oncological clearance.</p>	<p>During a radical neck dissection, the following critical extranodal anatomic structures are sacrificed:</p> <ol style="list-style-type: none"> 1. Sternocleidomastoid muscle (SCM) 2. Internal jugular vein (IJV) 3. Spinal accessory nerve (CN XI) <p>The contents and lymph nodes of levels I–V are also classically removed; however, the exact levels may vary slightly based on the location of the primary malignancy and nodal levels involved.</p>
4. What elective neck dissection (END) levels are recommended to include for a patient with cT2N0M0 right lateral tongue squamous cell carcinoma?	<p>For a patient with a cT2N0M0 right lateral tongue squamous cell carcinoma, the recommended END levels typically include the following:</p> <ol style="list-style-type: none"> 1. Level I: This includes the submental lymph nodes 2. Level II: This includes the upper jugular lymph nodes 3. Level III: This includes the middle jugular lymph nodes 4. Level IV: This includes the lower jugular lymph nodes 5. Level V: This includes the posterior triangle lymph nodes. <p>END aims to remove lymph nodes in the neck that are at risk of containing micrometastases, even if they are not clinically palpable. For a cT2N0M0 tumor, there is no evidence of regional lymph node involvement (N0), but END is performed to assess and clear the regional lymph nodes. The levels mentioned above cover the lymphatic drainage pathways relevant to the primary tumor site in the lateral tongue.</p>	<p>Based upon National Comprehensive Cancer Network (NCCN) guidelines, for a patient with cT2N0M0 right lateral tongue squamous cell carcinoma, the recommended END levels typically include at least levels I–III. Surgeons may also elect to perform a level I–IV selective neck dissection given the risk of occult nodal disease or skip metastases that may involve level IV. In addition, level IV may also be included in the event a free flap reconstruction is performed. The risks and benefits of this decision should be carefully considered and weighed against various individual patient and clinicopathologic features.</p>
5. When is an END indicated for a cT1N0M0 right lateral tongue squamous cell carcinoma based on depth of invasion (DOI)?	<p>In cases of cT1N0M0 right lateral tongue squamous cell carcinoma, an END is typically indicated based on the DOI. Generally, END is recommended if the DOI exceeds 4 millimeters (mm) or if there are other adverse features such as lymphovascular invasion, perineural invasion, or poorly differentiated histology.</p> <p>The rationale behind this recommendation is that deeper invasion increases the risk of occult metastasis in the regional lymph nodes. Therefore, even if there are no clinically palpable lymph nodes (cN0), the presence of adverse</p>	<p>The decision to perform an END should be based on the risk of occult nodal metastasis in the nodal basin. Based upon National Comprehensive Cancer Network (NCCN) guidelines, for a patient with cT1N0M0 right lateral tongue squamous cell carcinoma, an END is typically recommended for tumors with a DOI > 3 mm. For a DOI of < 2 mm, an END may be indicated</p>

(Cont'd...)

Table A1. (Continued)

Question	ChatGPT's answer	Surgeon's answer
	features, particularly DOI>4 mm, suggests a higher likelihood of occult metastasis, prompting the need for END to thoroughly evaluate and clear the regional lymph nodes.	<p>in highly select circumstances. For a DOI of 2–4 mm, careful clinical judgement to perform an END must be exercised; the surgeon may consider various clinicopathologic features (e.g., degree of clinical suspicion, patient age, reliability of follow-up for surveillance, etc.) in making this decision.</p> <p>Moreover, sentinel lymph node (SLN) biopsy is an alternative to END for the identification of occult nodal disease in early-stage (cT1 or cT2) oral cavity squamous cell carcinoma. This may inform the need for a completion or therapeutic neck dissection in the event of pathologically confirmed nodal disease or in the event that SLN biopsy is unsuccessful.</p>