AMERICAN ACADEMY OF OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

Otolaryngology–Head and Neck Surgery
2023, Vol. 00(00) 1–4

DOI: 10.1002/ohn.601
http://otojournal.org
WILEY

Using Generative Artificial Intelligence in the Production and Dissemination of Innovation in Otolaryngology—Ethical Considerations

Carolyn Jane Khoury, BA[1], Necati Enver, MD, FEBORL-HNS[2]*, Alberto Paderno, MD, PhD[3]*, Emanuele Ratti, PhD[4]*, and Anaïs Rameau, MD, MPhil, MS FACS[5]*

Keywords

generative AI, otolaryngology

Received August 8, 2023; accepted November 2, 2023.

Hypothetical Case Presentation

A team of otolaryngologists is studying the efficacy of early detection strategies in patients affected by idiopathic subglottic stenosis. They use an interactive artificial intelligence (AI) chatbot to prompt patients to report symptoms from home to remotely monitor the progression of stenosis. The chatbot prompts patients to provide feedback on their satisfaction with remote monitoring. Primary endpoints include extent of stenosis at evaluation (noted by flexible laryngoscopy in the office), number of unnecessary visits, and frequency of surgical interventions to relieve stenosis, such as CO_2-laser use, balloon dilation, or laryngotracheal resection. Secondary endpoints include oxygen saturation, spirometry score, pulse, and patient-reported outcomes. Because idiopathic subglottic stenosis is rare, the study did not have as large of a data set as they had hoped. Therefore, the team augments their existing data set via a generative adversarial network (GAN) trained on their existing data. The team reports significant improvement in the measured endpoints, with marked decrease in unnecessary visits and surgeries. The team has become occupied with clinical duties and does not have time to compose the manuscript. Moreover, they have a history of submitting poorly written manuscripts. Therefore, they decide to use generative AI to aid in manuscript production.

Point

Generative AI must not be used in otolaryngological research and manuscript production.

A Brief Introduction to Generative AI

To use generative AI technologies, users provide a prompt to the algorithm. "Summarize existing literature on idiopathic subglottic stenosis," for instance. The algorithm, equipped with an inconceivably large pool of data to draw upon, uses deep-learning networks to generate statistically probable outputs based on the initial prompt.[1] Large language models (LLMs) such as OpenAI's ChatGPT, generate highly probable text in the form of syntactically correct sentences.[1] In this case, the output would be a literature review synthesizing existing literature on idiopathic subglottic stenosis. GANs, another form of generative AI, augment data sets by generating synthetic instances from available data, and can also synthesize artificial images.

Lack of Pragmatism

Generative AI has no concept of reality, knowledge, or truth: simply data inputs and outputs. Biomedical ethicist Dr. David Magnus applies concepts from epistemology and linguistics to appraise the nature of AI-generated language, claiming "AI doesn't understand the pragmatics of language."[2] Pragmatics emphasizes the context in which we use language—speaker, audience, intent—in deriving meaning from language.[2] Generative AI lacks pragmatism because it lacks the context implicit in the expertise, knowledge, and realities of the practice of otolaryngology—the clinical context that informs and

[1]Sean Parker Institute for the Voice, Weill Cornell Medicine, New York, New York, USA
[2]Department of Otolaryngology–Head and Neck Surgery, Marmara University Faculty of Medicine, Istanbul, Turkey
[3]Otorhinolaryngology Unit, IRCCS Humanitas Research Hospital, Rozzano, Milan, Italy
[4]Department of Philosophy, University of Bristol, Bristol, UK
[5]Sean Parker Institute for the Voice, Department of Otolaryngology–Head and Neck Surgery, Weill Cornell Medical College, New York, NY, USA
*These authors contributed equally to this article and are considered co-last authors.

Corresponding Author:
Anaïs Rameau, MD, MPhil, MS FACS, Sean Parker Institute for the Voice, Department of Otolaryngology–Head and Neck Surgery, Weill Cornell Medical College, New York, NY, USA.
Email: anr2783@med.cornell.edu

inspires a line of research. This raises concerns about how this technology grapples with concepts and data without the context crucial to deriving meaning and truth from it, making generative AI not only fallible, but also unable to reliably appraise the data it processes. A dangerous conflation occurs when this technology's impressive syntactical predictive power generates natural language convincing enough to be taken as intelligence. Therefore, one has a duty to apply the pragmatic context of one's otolaryngological expertise to the semantic prowess of generative AI in order to uphold the integrity and clinical relevance of one's research.

Generative AI Is Not Infallible

Generative AI may produce factual and mathematical errors, generating false information that it presents as truth, referred to as "hallucination."[3] It generates answers based not on correctness, but on the most probable output given the training data.[3] Currently, hallucinations cannot be consistently prevented or identified, severely impacting the reliability of concepts expressed by LLMs. Furthermore, the model's output relies heavily on its input prompt. Similar instructions may generate markedly heterogeneous outputs expressing different concepts and even contradictory logic flows. Specifically, researchers might inadvertently induce a framing effect in the prompt based on their desired outcome, leading to conclusions based not on data, but on their expectations.[4]

In the hypothetical case presentation, a GAN was used to augment the data set based on actual collected data. But what if the data were inaccurate or misleading? Augmentation with synthetic data should follow strict criteria and reporting standards and should not be used to increase the statistical significance of key endpoints with an insufficient sample size. Otherwise, incorrect information could be taken as truth, thus exposing patients to a potentially harmful product. Moreover, such erroneous conclusions could inspire further research, consequently increasing the extent of harm by creating a foundation of inquiry based on misinformation. The algorithm does not have the context—the pragmatism—with which to appraise its outputs; it cannot distinguish between truth and falsity. The consequences can clearly be detrimental, ultimately violating the bioethical principle of nonmaleficence. Therefore, to uphold the integrity of one's research, we have an obligation to not use or publish conclusions drawn by generative AI.

Input Bias and Epistemic Injustice

Generative AI incorporates bias from the data it is trained upon, thus perpetuating existing biases. One consequence is an AI-content-loop.[5] As more AI-generated content populates the pool of existing literature, generative AI algorithms will ultimately come to be trained by their own outputs. This is posited to lead to "model collapse," in which AI-generated data corrupts the training input for the next generation of AI algorithms, resulting in a body of literature representing an echo chamber of AI-generated outputs with an increasingly distorted picture of reality. If otolaryngology's literature is shaped by this AI-polluted echo chamber, it will shape how readers perceive the state of the field, thus inspiring innovators to pursue routes of inquiry influenced by a corrupt knowledge base skewed by the algorithms' assessment of what is most probable. Another issue related to input bias is plagiarism.[3] Take the aforementioned literature review on subglottic stenosis. The program could borrow sentences and ideas directly from the input literature without proper citations. This runs the risk of appropriating colleagues' intellectual property. One could take this further and argue that using any LLM to write text may be considered plagiarism since the work is fundamentally not one's own. Finally, sharing data with generative AI may infringe on patient privacy to the benefit of generative AI companies, and the ramifications of data sharing need to be clearly discussed with patients via the informed consent process.

Medical Image Falsification

A burgeoning application of generative AI in otolaryngology is the production of synthetic medical images. Conditional GANs (cGANs) and text-to-image platforms such as DALL-E, Stable Diffusion, and Midjouney quickly produce images with profound accuracy and realism. The generation of realistic biomedical images with AI, such as Western blots, immunohistochemical stains, magnetic resonance imaging, pictures of surgical outcomes, and so on, provides ample opportunity for the falsification of medical images to support a claim. This deceit can directly harm patients secondary to medical knowledge based on unfounded and false data. Even more frightening, it threatens scientific inquiry at large. A central component of many scientific methodologies is observation—using our senses, such as sight—to understand and raise questions about the world around us. The ease with which images can now be falsified threatens our ability to trust what we observe. Such deepfakes may also undermine fraud detection through the peer-review process.

Counterpoint

The use of generative AI is acceptable because it can accelerate, augment, and inspire advancements in otolaryngology, and their dissemination as manuscripts.

Increased Efficiency of Clinical Research

The manifold applications of generative AI throughout the scientific method hold exciting implications for the advancement of otolaryngology, and biomedicine at large. The most direct benefit of generative AI to manuscript production is its ability to save time on tasks that do not require creativity or scientific reflection.[3] Academic otolaryngologists balance many responsibilities including clinical and surgical responsibilities, teaching, and personal commitments, along with participation in research projects. Increasing the speed with which one can produce and report their innovations as manuscripts is, therefore, a significant boon. Returning to our case presentation, if the researchers input relevant papers on idiopathic subglottic stenosis into an LLM, they could prompt the program to create a literature review in less than a minute. This efficiency allows investigators to spend less time belaboring over monotonous and mechanical aspects of manuscript production, and more time on the ingenuity and creativity of innovation. Additionally, generative AI can suggest relevant literature to writers that they may not have previously encountered, thus enhancing their understanding of the field and perhaps inspiring different approaches to their study.

Furthermore, LLMs open the door to investigators who find manuscript-writing to be a challenge precluding them from contributing their thoughts to the community, such as those for whom English is a second language. AI-assisted academic writing can reduce the disparity between native English speakers and those using English as a second language, making the scientific landscape a more inclusive and equitable forum. Moreover, LLMs can help otolaryngologists enhance their academic writing skills. Writers can prompt LLMs to not only edit manuscripts, but also identify points of weakness in their writing style such that writers may improve upon them in the future.

The increases in efficiency offered by generative AI extend to clinical data collection.[6] As exemplified in the case presentation, generative AI chatbots can interface with patients or caregivers to collect patient information, or even administer questionnaires to obtain patient-reported outcome measures.

Augmentation of Human Ingenuity

Generative AI can open new avenues of inquiry and creativity that would never have been considered given the current state of otolaryngology. The technology's lack of otolaryngological context allows it to impartially connect concepts from different fields to present outside-the-box ideas that experts entrenched in the field may not have considered, thus combatting expertise bias.[7] This exposes otolaryngologists to new ways of thought which they can synergize with their otolaryngological expertise to create novel solutions to the field's current dilemmas. For instance, in our clinical scenario, generative AI may provide insights from medical literature beyond otolaryngology that could be relevant to the management of subglottic stenosis.

Additionally, LLMs can appraise a study's practicality.[7] When prompted with detailed information about the goals of a study, this technology can assess the study's feasibility and impact; propose methodologies and workflows; anticipate unforeseen risks; and even provide real-time reports of project status, suggesting changes to study design accordingly.[7] In this way, generative AI maximizes the productive capacity of a project and provides the logistical scaffold upon which the innovators of otolaryngology can bring life-enhancing and life-saving ideas to fruition.

Enhanced Production, Utility, and Analysis of Medical Images

Image-generating-AI has been employed to create synthetic images used to train automated diagnostic tools. One group employed cGANs to create synthetic images of normal and abnormal tympanic membranes, which they use to train a machine-learning diagnostic tool for otitis media.[8] Not only did they produce incredibly high-quality images of the tympanic membrane, but they also found that diagnostic algorithms trained on the synthetic images were better able to distinguish between normal and abnormal tympanic membranes in real-world patients than those trained on real, nonsynthetic images. Generating synthetic medical images has been particularly applicable to rare pathologies suffering from a scarcity of raw data. In such scenarios, synthetic data can be utilized to feed the model with fundamental disease features, enabling it to establish a rudimentary domain representation of the pathology. This approach obviates the need for an extensive collection of images, thus offering a potential solution to the inherent data limitation issue.

Potential Solutions to the Methodological and Ethical Drawbacks

The advantages of generative AI will not materialize without effort from the otolaryngology community to create benchmarks and standards to facilitate their correct implementation. The drawbacks of LLMs in biomedicine derive from using these tools without adapting them to the specificities of otolaryngology and its practices.[9] Moreover, standards for how LLMs' parameters are fine-tuned to their specific tasks must be established. Finally, task-specific labeled data must be generated, curated, and standardized accordingly. There is no portability of LLMs across different contexts without careful adaptation. To ensure critical oversight and moral integrity in the use of generative AI, its regulation in medical research ought to be codified in law.[10] A "total product lifecycle approach" should be used to regulate this technology as it is rapidly evolving.[10]

Generative AI has undoubtedly exciting prospects that will propel otolaryngology into a new era of exploration. The new avenues of exploration it inspires, coupled with the ease with which it allows us to disseminate novel findings through efficiently written manuscripts, will pave the way for otolaryngologists to heal more patients in more and better ways. Still, we must keep in mind that with generative AI's great probabilistic power must come great responsibility from investigators and regulators. We must approach AI-generated outputs with the utmost caution to prevent undue harm and preserve the immense healing capacity of this exciting technology.

Author Contributions

Carolyn Jane Khoury, study design, literature review, and manuscript preparation; **Necati Enver**, study design, manuscript revision; **Alberto Paderno**, study design, manuscript revision; **Emanuele Ratti**, study design and manuscript revision; **Anaïs Rameau**, conception, study design, case presentation, manuscript revision; Weill Cornell Medicine IRB approval or exemption was not sought for this study, since no data collection was performed, and contents of this manuscript are based on literature review and reflections by the authors.

Disclosures

Competing interests: Anaïs Rameau owns equity of Perceptron Health Inc. Anaïs Rameau is medical advisor for Savorease, Inc. Alberto Paderno owns equity of CareForMe Ltd. Alberto Paderno is medical advisor for HealthX Ltd.

Funding source: 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. Emanuele Ratti was supported in part by an award from the Notre Dame-IBM Tech Ethics Lab (number 262812UB) and Innovate UK and Horizon Europe (project id 10063119). Such supports do not constitute endorsement by the sponsor of the views expressed in this publication.

ORCID iD

Carolyn Jane Khoury http://orcid.org/0009-0007-4558-3907
Necati Enver https://orcid.org/0000-0002-3161-8810
Alberto Paderno https://orcid.org/0000-0002-1621-2142
Emanuele Ratti https://orcid.org/0000-0003-1409-8240
Anaïs Rameau https://orcid.org/0000-0003-1543-2634

References

1. Korzynski P, Mazurek G, Altmann A, et al. Generative artificial intelligence as a new context for management theories: analysis of chatgpt. *Cent Eur Manag J*. 2023; 31(1):3-13.
2. Link K, Magnus D. Ethical Issues Raised by LLM's and Generative Models in Stanford University Artificial Intelligence in Medicine & Imaging (AIMI) Symposium 2023. June 8, 2023.
3. Sallam M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. *Healthcare*. 2023;11(6):887.
4. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. Bender EM, Gebru T, McMillan-Major A, Shmitchell A. On the dangers of stochastic parrots: can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT), March 3-10, 2021, Virtual Event, Canada. 2021:610-623.
5. Jin C, Tan F, Jiang S. Generative adversarial network technologies and applications in computer vision. computational intelligence and neuroscience. *Comput Intell Neurosci*. 2020;2020:1-17.
6. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial Intelligence and machine learning: systematic review. *JMIR Cancer*. 2021;7(4):e27850.
7. Eapen TT, Finkenstadt DJ, Folk J, Venkataswamy L. How generative AI can augment human creativity. *Harvard Business Review*. June 13, 2023. Accessed June 26, 2023. https://hbr.org/2023/07/how-generative-ai-can-augment-human-creativity
8. Suresh K, Cohen MS, Hartnick CJ, Bartholomew RA, Lee DJ, Crowson MG. Making use of artificial intelligence–generated synthetic tympanic membrane images. *JAMA Otolaryngol Head Neck Surg*. 2023;149(6):555-556.
9. Bommasani R, Hudson DA, Adeli E, et al. On the opportunities and risks of foundation models. *arXiv preprint*. 2021:2108.07258.
10. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. 2023;6(1):120.