AMERICAN ACADEMY OF OTOLARYNGOLOGY–HEAD AND NECK SURGERY FOUNDATION

Scientific Briefing

Otolaryngology–Head and Neck Surgery
2025, Vol. 173(4) 1024–1027

DOI: 10.1002/ohn.1335
http://otojournal.org
WILEY

Expanding the Capacity of General Practitioners in Sub-Saharan Africa With Artificial Intelligence

Jerome R. Lechien, MD, PhD, MS[1,2,3,4]

Abstract

Low- and middle-income countries face substantial healthcare delivery challenges due to specialist shortages. This study assessed POE (Q-Exp-Claude-3.5-2kk), a multimodal large language model, for diagnostic accuracy and clinical utility in otolaryngology across 63 consecutive patients in rural Kenya. Patients underwent evaluation by both an otolaryngologist and primary care practitioner (PCP), with clinical data and micro-otoscopic images submitted to POE's interface. POE's primary diagnoses demonstrated 79.4% concordance with otolaryngologist findings, while management recommendations aligned in 96.8% of cases. Differential diagnoses were judged plausible and correct in 73.0% of cases. POE demonstrated significantly superior diagnostic accuracy compared to PCPs (79.4% vs 50.8%, $P = .001$). Among cases where PCPs failed to establish a diagnosis, POE correctly identified 61.3%. With examination images, POE accurately identified conditions as primary (50.0%) or potential (35.7%) diagnoses, demonstrating value as a clinical decision support tool in resource-limited settings.

Keywords

Africa, artificial intelligence, head neck surgery, humanitarian, low, middle-income, mission, otolaryngology, otorhinolaryngology, outreach, Sub-Saharan

Received March 4, 2025; accepted May 31, 2025.

Most Sub-Saharan Africa has limited otolaryngology and audiology capacity with 1.2 otolaryngologists and 0.12 audiologists per 1,000,000 inhabitants.[1] However, nearly 90% of people with hearing disorders live in low- and middle-income countries,[2] and predictive studies suggested that by 2030, the number of patients with hearing disorders and cancers will increase in developing countries.[2] Despite limited access to healthcare resources, the population has access to a substantial telecommunications infrastructure with 88% of Kenyan adults with a cellphone, 81% having access to the internet.[2]

This brief communication aimed to evaluate POE® (Quora Inc.), a multimodal large language model platform, for diagnostic accuracy, clinical consistency, and usefulness as an adjunctive tool for otolaryngological care in resource-limited rural Kenyan settings.

Methods

Patients and Setting

Demographics and clinical findings of consecutive outpatients consulting in the Iten facility for otolaryngological conditions were prospectively collected from December 14, 2024, to January 3, 2025, in a humanitarian outreach by the author of the paper. The Iten county hospital is a facility organized into various departments, offering services beyond primary care. It operates with limited infrastructure and resources intermediate between a dispensary and a full-service general hospital. Ear, nose, and throat (ENT) consultation is assumed by a primary care practitioner (PCP) without basic equipment (nasofibroscopy, ear microscope, tympanometry device, or otoacoustic emissions). The local ethics committee approved the study protocol (IRCH-20251). Patients consented to participate.

Outcomes

The consultation was conducted by the PCP of the ENT office (Iten hospital), and a board-certified otolaryngologist in humanitarian outreach. Practitioners used a new cheap ($30) HD-otoscopic device, which was developed for ear wax removal, for performing the anterior nasal cavity and otoscopic examinations (Bebird R1 earscope, Shenzhen Technology-Co). Real-time visualization of the external ear duct, tympanic membrane, or nasal cavities is achieved through a dedicated smartphone application

[1]Phonetics and Phonology Laboratory (UMR 7018 CNRS, Université Sorbonne Nouvelle/Paris 3), Department of Otolaryngology–Head and Neck Surgery, Foch Hospital, School of Medicine, Paris Saclay University, Paris, France
[2]Department of Surgery, UMONS Research Institute for Health Sciences and Technology, University of Mons, Mons, Belgium
[3]Department of Otolaryngology, Elsan Polyclinic de l'Atlantique, Poitiers, France
[4]Department of Otolaryngology–Head and Neck Surgery, CHU Saint-Pierre, Brussels, Belgium

Corresponding author:
Jerome R. Lechien, MD, PhD, MS, Department of Otolaryngology–Head and Neck Surgery, Elsan Hospital, Poitiers, France.
Email: Jerome.Lechien@umons.ac.be

(Suear® for iOS and Android) via Wi-Fi connectivity (Figure 1).

The clinical information, including the video/pictures of the clinical examination, was entered into the API of POE® through a standardized sentence according to recent studies[3]: "I am physician in a Humanitarian setting with limited resources. My patient is a female/male of … years with the following symptoms and findings: … The medical/surgical history include: … The current treatment is: … What are your primary and differential diagnoses?; What are your additional examinations to find the diagnosis?; and What are your treatment(s) for the primary diagnosis?" For clinical images, patient identity was protected. When facial images were necessary, we placed a black bar across the eyes and focused the image on the relevant semiological findings. In this study, Q-Exp-Claude-3.5-2kk was used for generating clinical outputs, which were compared with those of the PCP by the otolaryngologist. The performance of POE-Q-Exp-Claude-3.5-2kk, especially for clinical cases with uncertain diagnosis and management from the PCP, was evaluated with the artificial intelligence performance instrument (AIPI).[3] AIPI includes nine items evaluating the management of a clinical case by AI/large language model (LLM) (Supplemental Figure S1, available online). AIPI is subdivided into the four following subscores associated with common items: patient feature score (/6), diagnosis score (/7), additional examination score (/5), and treatment score (/3). The scoring of items was defined to be as objective as possible, avoiding the use of Likert scale. AIPI score ranges from 0 to 20 with a score of 20 indicating excellent clinical case management by the AI/LLM. AIPI provides a comprehensive approach to clinical cases, intended for use not only in otolaryngology but also in general medicine and surgery. The difficulty of clinical cases was evaluated by the otolaryngologist with a 5-point Likert scale ranging from 1 (no difficulty) to 5 (very difficult case).

Statistical Methods

Statistical analyses were performed using the Statistical Package for the Social Sciences for Windows (SPSS version 29.0; IBM Corp). AIPI scores were comparatively analyzed between cases with and without clinical image uploads, and between cases stratified by PCP diagnostic status (established diagnosis vs no established diagnosis). Bivariate analysis was conducted between consultation outcomes.

Results

Sixty-three consecutive patients were included. Demographics and diagnosis findings are reported in **Table 1**. There were 37 children (58.7%) and 26 females (41.3%). The mean age was 24.2 ± 25.2 years. The mean level of difficulty was 1.66 ± 1.03. Fifty-nine patients (93.7%) did not report comorbidities. Suspected allergic rhinitis, adenotonsillar hypertrophy, reflux, and cholesteatoma were the most prevalent primary diagnoses according to the otolaryngologists. The primary diagnoses generated by POE-Q-Exp-Claude-3.5-2kk demonstrated concordance with otolaryngologist findings in 50/63 cases (79.4%), whereas management recommendations aligned in 61 cases (96.8%). Differential diagnoses were judged as plausible and correct in 46 cases (73.0%). Among the discordant cases (n = 17), POE correctly included the confirmed diagnosis within its differential considerations in 16 outputs.

Primary care physicians either failed to establish a diagnosis or provided discordant assessments compared to otolaryngologist determinations in 49.2% of cases (31/63). POE-Q-Exp-Claude-3.5-2kk demonstrated superior

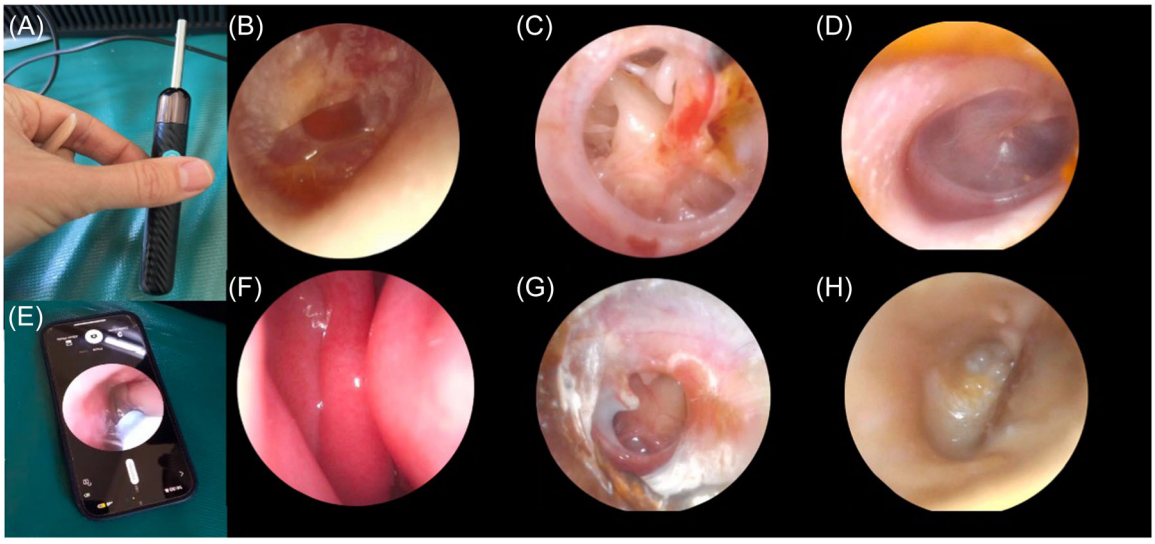

Figure 1. Ear device images. The device incorporates a high-precision optical system featuring a 3-megapixel sensor, six LED illumination sources, and 1080P resolution capability. Real-time visualization of images is achieved through a smartphone application (Suear®).

Table 1. Characteristics of Patients

Demographics	Adults (N = 26)	Children (N = 37)	Total (N, %)
Age (mean, SD)	49.1 ± 18.7	5.1 ± 3.1	24.2 ± 25.2
Gender (N, %)			
Females	14 (53.8)	12 (32.4)	26 (41.3)
Males	12 (46.2)	25 (67.6)	37 (58.7)
Primary otolaryngologist's diagnoses			
Suspected allergic or nonallergic rhinitis	4 (15.4)	5 (13.5)	9 (14.3)
Adenoid/tonsil hypertrophy and apnea	0 (0)	8 (21.6)	8 (12.7)
Suspected LPRD (RSS > 13)	4 (15.4)	1 (2.7)	5 (7.9)
Suspected GERD (Lyon consensus) with LPRD (RSS > 13)	4 (15.4)	1 (2.7)	5 (7.9)
Chronic otitis media with cholesteatoma	3 (11.5)	2 (5.4)	5 (7.9)
Acute pharyngitis (virus, foreign body irritation)	3 (11.5)	1 (2.7)	4 (6.3)
Recurrent tonsillitis	0 (0)	4 (10.8)	4 (6.3)
Chronic otitis media with suppuration (no complication)	0 (0)	4 (10.8)	4 (6.3)
Adenoid hypertrophy with/without adenoiditis	0 (0)	3 (8.1)	3 (4.8)
Snoring without apnea related to nasal deviation	2 (7.7)	0 (0)	2 (3.2)
Epistaxis	1 (3.8)	1 (2.7)	2 (3.2)
Chronic adhesive otitis media	1 (3.8)	1 (2.7)	2 (3.2)
Thyroid goiter	1 (3.8)	0 (0)	1 (1.6)
Idiopathic dysphonia (suspicion of vocal cord cancer)	1 (3.8)	0 (0)	1 (1.6)
Laryngotracheitis (acute)	1 (3.8)	0 (0)	1 (1.6)
Eustachian tube dysfunction	1 (3.8)	0 (0)	1 (1.6)
Diabetes mellitus	1 (3.8)	0 (0)	1 (1.6)
Nasal foreign body	0 (0)	1 (2.7)	1 (1.6)
Complicated maxillary rhinosinusitis (acute)	0 (0)	1 (2.7)	1 (1.6)
Ear wax	0 (0)	1 (2.7)	1 (1.6)
External otitis (acute)	0 (0)	1 (2.7)	1 (1.6)
External ear duct eczema	0 (0)	1 (2.7)	1 (1.6)
Short tongue-ties	0 (0)	1 (2.7)	1 (1.6)
Otoscopic device usefulness			
Clinical usefulness (N, %)	6 (23.1)	21 (56.8)	27 (42.9)[a]
Diagnostic necessity (N, %)	6 (23.1)	8 (21.6)	14 (22.2)
AIPI scores with images (mean, SD)	-	-	16.4 ± 3.8[b]
AIPI scores without images (mean, SD)	-	-	18.3 ± 2.9

Abbreviations: AIPI, artificial intelligence performance instrument; GERD, gastroesophageal reflux disease; LPRD, laryngopharyngeal reflux disease; N, number; RSS, reflux symptom score; SD, standard deviation.

[a]Statistical significant differences.

[b]*P*-value comparing AIPI scores with and without images.

diagnostic accuracy compared to PCPs (79.4% [50/63] vs 50.8% [32/63], $P = .001$) in establishing primary diagnoses (**Table 2**). Among cases where PCPs failed to establish the diagnosis, POE-Q-Exp-Claude-3.5-2kk successfully identified the primary diagnosis in 61.3% (19/31). Cases with PCP-established diagnoses (n = 32) demonstrated higher AIPI diagnostic subscores compared to undiagnosed cases (n = 31).

Evaluation of the otoscopic device efficacy reported significant clinical usefulness in 27 cases (42.9%, 21 children vs 6 adults) and absolute diagnostic necessity in 14 cases (22.2%, 8 children vs 6 adults). Some clinical pictures are reported in **Figure 1**. The otoscopic device demonstrated significantly higher usefulness in the pediatric cohort compared with the adult group ($P = .014$). When provided with clinical examination images, POE-Q-Exp-Claude-3.5-2kk correctly identified the condition as either primary (n = 7/14, 50.0%) or potential (n = 5/14, 35.7%) diagnoses (**Table 1**). The mean AIPI score of picture-based cases (16.4 ± 3.8) was significantly lower than the score for cases without a clinical picture (18.3 ± 2.9; $P = .011$).

The correlation study reported a significant association between the level of difficulty and age ($r_s = 0.454$; $P = .001$). The AIPI score was not influenced by the level of difficulty.

Discussion

This preliminary study demonstrates the usefulness and accuracy of using AI-based tuned LLM/platform as an

Table 2. Comparison Between POE and Primary Care Practitioner

Parameters	All patients (N = 63)		*P*-value
	POE	Primary care	
Differential diagnosis			
Complete or incomplete but plausible	46 (73.0)	32 (50.8)	.015
Incomplete, not plausible, or absent	17 (27.0)	31 (50.2)	
Primary diagnosis			
Correct or plausible	50 (79.4)	32 (50.8)	.001
Not plausible or absent	13 (20.6)	31 (50.2)	
Management recommendation consistent with otolaryngologist	61 (96.8)	35 (55.6)	.001
AIPI score for POE findings			
Patient feature scores (/6: mean, SD)	5.7 ± 0.8	-	-
Diagnosis subscore (/7: mean, SD)	6.2 ± 1.3	-	-
Additional Examination score (/4: mean, SD)	3.4 ± 0.9	-	-
Treatment score (/1: mean, SD)	2.5 ± 0.8	-	-
AIPI total score	17.9 ± 3.1	-	-

Abbreviation: AIPI, artificial intelligence performance instrument.

adjunctive tool in Sub-Saharan Africa facilities affected by the shortage of specialist practitioners. The accuracy of LLMs, including ChatGPT-4o and Claude-3.5-Sonnet, was evaluated in previous studies,[3-5] demonstrating an accuracy ranging from 38% to 86% of cases for the primary diagnosis of ChatGPT-4o,[4] whereas the tuned Q-Exp-Claude-3.5-2kk reached 54.3% for very rare conditions.[5] Combining e-Health devices, such as the otoscopic device, with e-Health applications and platforms in low- and moderate-income countries with sufficient telecommunications infrastructure can provide an opportunity for improving the care delivery in rural regions. Although this study is unique in the otolaryngology literature, the low number of patients, the single otolaryngologist assessment, and the lack of materials for controlling the final diagnosis are the primary limitations of the study and, therefore, need to be addressed in future large cohort studies.

Conclusion

The implementation of integrated LLM platforms in low- and middle-income countries demonstrates usefulness as clinical decision support tools, augmenting PCP diagnostic and therapeutic capabilities.

Acknowledgments

The author thanks B. Johnson for the proofreading.

Author Contributions

Jerome R. Lechien, Design; acquisition of data; data analysis and interpretation; drafting; final approval; accountability for the work; final approval of the version to be published; agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Disclosures

Competing interests: The author has no conflict of interest.

Funding source: The author has no financial disclosures.

Supplemental Material

Additional supporting information is available in the online version of the article.

References

1. Mulwafu W, Fagan JJ, Mukara KB, Ibekwe TS-. ENT outreach in Africa: rules of engagement. *OTO Open*. 2018;2(2):2473974X18777220. doi:10.1177/2473974X18777220
2. Jayawardena ADL, Kahue CN, Cummins SM, Netterville JL. Expanding the capacity of otolaryngologists in Kenya through mobile technology. *OTO Open*. 2018;2(1):2473974X18766824. doi:10.1177/2473974X18766824
3. Maniaci A, Chiesa-Estomba CM, Lechien JR. ChatGPT-4 consistency in interpreting laryngeal clinical images of common lesions and disorders. *Otolaryngol Head Neck Surg*. 2024;171(4):1106-1113. doi:10.1002/ohn.897
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. Lechien JR. Large language models as adjunctive tools for diagnosing rare disease in otolaryngology: a controlled study. University of Mons Grands Rounds; 2025.