

AI CHATBOTS IN THE RETINA CLINIC

Large language models are not coming for your job.
But they may redefine what your job looks like.

By Fares Antaki, MDCM



AI has been in ophthalmology for years, and most retina specialists are familiar with AI as it relates to image analysis. Algorithms that detect diabetic retinopathy (DR) from fundus photographs or segment fluid on OCT have received the most attention and, in the case of autonomous DR screening, regulatory clearance.^{1,2} However, a second wave of AI is reaching the clinic, and this version can read, write, and reason with text. These large language models (LLMs) power the chatbots that are quietly changing how physicians work.³

WHAT ARE LARGE-LANGUAGE MODELS?

LLMs are a class of AI designed to understand and generate human language. The technology behind products such as ChatGPT (OpenAI), Gemini (Google), and Claude (Anthropic) is conceptually straightforward. During training, a model is shown massive amounts of text and learns to predict the next word in a sequence.⁴ Consider the phrase attributed to Heraclitus: “The only constant in life is change.”⁵ A model in training would see “The only constant in life is...” and learn to predict “change.” This process is repeated billions of times across the entirety of the internet, books, and encyclopedias. The result is a system with a broad representation of human knowledge.

These models can hold multi-turn conversations, interpret nuanced queries, generate structured outputs, and adapt tone and style to context. However, because these inputs contain misinformation, errors, and bias, LLMs must undergo a process called *alignment* to ensure they are helpful, honest, and harmless (3H), aligning with human values.⁶ In practice, human evaluators grade thousands of model outputs, and these preferences are used to fine-tune the

model toward 3H behavior.⁷ Despite alignment, LLMs can still produce confident but incorrect and potentially harmful statements—a limitation critical for clinical applications.

OPHTHALMIC KNOWLEDGE

LLMs have shown impressive performance in medicine, specifically on ophthalmology knowledge assessments.⁸ In our research, we tested successive generations of GPT models on a 260-question multiple-choice dataset licensed from the AAO, comparable with the Ophthalmic Knowledge Assessment Program. Each iteration of the AI performed better: GPT-3.5 scored 55.8%, GPT-4 achieved 75.8%, and GPT-5 scored 96.5%.⁹⁻¹¹

Multiple groups have now demonstrated LLM competence not only on ophthalmic board-style questions

KEY TAKEAWAYS

- ▶ Large language models have embedded ophthalmic knowledge that is improving rapidly with each model generation.
- ▶ Practical applications for retina specialists span two categories: workflow improvement (electronic health record documentation, report generation) and clinical assistance (triage, decision support).
- ▶ Vibe coding allows retina specialists without programming experience to build custom clinical tools using large-language models.



RETINA SPECIALISTS ADOPTING THESE TOOLS MUST DO SO WITH APPROPRIATE CAUTION, TREATING LLM OUTPUTS AS ADVISORY RATHER THAN AUTHORITATIVE.

but also as a clinical assistance tool and for patient education.^{12,13} Each model generation embeds deeper medical knowledge and demonstrates stronger clinical reasoning. Most recently, a new class of reasoning LLMs can break down complex clinical problems into step-by-step chains of thought before producing an answer, further narrowing the gap with expert-level performance.^{14,15} Thus, LLMs have accumulated enough ophthalmic knowledge to serve as useful tools in the hands of a trained clinician.

PRACTICAL APPLICATIONS IN RETINA

Applying LLM chatbots in the retina clinic falls into two broad categories: workflow improvement and clinical assistance tools.¹⁶

In workflow improvement, the most immediate use case is documentation. AI scribes that listen to patient encounters and generate structured notes are already in clinical use.¹⁷ Chatbots can also draft referral letters, extract relevant data from lengthy records and summarize the chart, and prepare patient education materials.¹⁸ These tasks consume a significant portion of a retina specialist's day. Offloading them to an LLM does not replace clinical judgment; it reclaims time.

A more forward-looking application is clinical decision support. In daily life, consumers use chatbots to answer health questions. More than 230 million people globally ask health- and wellness-related questions on ChatGPT every week.¹⁹ LLMs can process unstructured patient descriptions, ask targeted follow-up questions, and generate a prioritized differential diagnosis and triage suggestions in real time. Consider the potential utility of a 24/7 triage chatbot where a patient describes acute symptoms after an intravitreal injection, and the system flags concern for endophthalmitis and directs them to the emergency department. Early-stage systems like this are being explored, although regulatory and liability frameworks remain undeveloped.

LLMs can also function as on-demand consultants during clinical decision making. A retina specialist uncertain about a rare presentation can describe the case to the chatbot and receive a differential diagnosis with supporting reasoning. In a recent study, our group partnered with Google to evaluate AMIE, a medically fine-tuned LLM based on Gemini, across 100 ophthalmology vignettes. After reviewing AMIE's output, clinicians improved their accuracy from 83.7% to 87.3%, revised their management plans in 34% of cases, and tended

to rank the correct diagnosis higher in their differential.²⁰ The LLM response is not a substitute for clinical expertise, but it can serve as a cognitive aid, particularly for trainees or in settings with limited specialty access.

VIBE CODING: BUILDING YOUR OWN TOOLS

Perhaps the most underappreciated application of LLMs is their capacity to turn clinicians into software developers. The term *vibe coding* refers to a software development approach where the user describes what they want in plain language and the chatbot writes the code.²¹

For example, I wanted to streamline referrals to my retina practice. I described the requirements to the chatbot in Replit (a vibe coding platform): Build a web application that allows optometrists and other doctors to submit referrals with patient demographics, visual acuity, IOP, examination findings, and suspected diagnosis. Within minutes, I had a functional patient intake form with Snellen notation placeholders, dropdown menus for common diagnoses, and an image upload feature for fundus photographs and OCTs. The entire process required no programming knowledge.

Every retina specialist has inefficiencies in their practice that could be addressed by a simple tool. The barrier to building that tool has effectively been removed.

LIMITATIONS AND GUARDRAILS

Enthusiasm for LLMs must be tempered by an honest assessment of their limitations. LLMs can generate statements that are fluent, internally consistent, and wrong, known as *confabulations* or *hallucinations*.²² In a clinical context, an incorrect differential or a fabricated management plan could have direct and serious consequences for patients. Any information generated by an LLM must be carefully verified by the clinician before acting on it.

Data privacy is another consideration. Entering protected health information into a commercial chatbot raises compliance concerns under HIPAA and equivalent regulations. Enterprise-grade and on-premises solutions exist, but most retina specialists interacting with chatbots today are using consumer-facing products. Awareness of what information is being shared and with whom is essential.

Finally, the medicolegal landscape is undefined. If a chatbot-assisted triage decision leads to a missed diagnosis, where does liability fall? These questions are being debated but have not been resolved. Retina specialists adopting these



tools must do so with appropriate caution, treating LLM outputs as advisory rather than authoritative.

TIME TO INTEGRATE

Chatbots will never replace retina specialists. The retina clinic requires procedural skill, clinical intuition built from pattern recognition over thousands of cases, and a patient-physician relationship that no AI can replicate. However, the technology is imperfect but improving at a rate outpacing most physicians' awareness of it. The question is no longer whether AI will play a role in the retina clinic; it's whether you will be the one shaping that role or reacting to it. ■

AI disclosure: Claude Haiku 4.5 was used for language editing. All content was reviewed, verified, and revised by the author, who assumes full responsibility for the accuracy and integrity of the manuscript.

1. De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med*. 2018;24(9):1342-1350. doi.org/10.1038/s41591-018-0107-6
2. Abràmoff MD, Lavin PT, Birch M, Shah N, Folk JC. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit Med*. 2018;1(1):39. doi.org/10.1038/s41746-018-0040-6
3. Betzler BK, Chen H, Cheng CY, et al. Large language models and their impact in ophthalmology. *Lancet Digit Health*. 2023;5(12):e917-e924. doi.org/10.1016/S2589-7500(23)00201-7
4. Chia MA, Antaki F, Zhou Y, Turner AW, Lee AY, Keane PA. Foundation models in ophthalmology. *Br J Ophthalmol*. 2024;108(10):1341-1348. doi.org/10.1136/bjo-2024-325459
5. Kolokythas A. Greek philosophers: stoics and health care. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2022;133(5):501. doi.org/10.1016/j.oooo.2022.01.019
6. Bai Y, Jones A, Ndousse K, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. Cornell University. April 12, 2022. Accessed June 9, 2026. arxiv.org/abs/2204.05862
7. Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback. Cornell University. March 4, 2022. Accessed June 9, 2026. arxiv.org/abs/2203.02155
8. Agnihotri AP, Nagel ID, Artiaga JCM, Guevarra MCB, Sosuan GMN, Kalaw FGP. Large language models in ophthalmology: A review of publications from top ophthalmology journals. *Ophthalmol Sci*. 2025;5(3):100681. doi.org/10.1016/j.xops.2024.100681
9. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. *Ophthalmol Sci*. 2023;3(4):100324. doi.org/10.1016/j.xops.2023.100324
10. Antaki F, Milad D, Chia MA, et al. Capabilities of GPT-4 in ophthalmology: an analysis of model entropy and progress towards human-level medical question answering. *Br J Ophthalmol*. 2024;108(10):1371-1378. doi.org/10.1136/bjo-2023-324438
11. Antaki F, Mikhail D, Milad D, et al. Performance of GPT-5 frontier models in ophthalmology question answering. Cornell University. August 14, 2025. Accessed June 9, 2026. arxiv.org/abs/2508.09956
12. Luo MJ, Bi S, Pang J, et al. A large language model digital patient system enhances ophthalmology history taking skills. *NPJ Digit Med*. 2025;8(1):502. doi.org/10.1038/s41746-025-01841-6
13. Chen X, Zhao Z, Zhang W, et al. EyeGPT for patient inquiries and medical education: Development and validation of an ophthalmology large language model. *J Med Internet Res*. 2024;26(1):e60063. doi.org/10.2196/60063
14. Srinivasan S, Ai X, Zou M, et al. Ophthalmological question answering and reasoning using OpenAI of vs other large language models. *JAMA Ophthalmol*. 2025;143(9):740-748. doi.org/10.1001/jamaophthalmol.2025.2413
15. Wang X, Xiong Z, Zou K, et al. Reasoning-driven large language models in medicine: opportunities, challenges, and the road ahead. *Lancet Digit Health*. 2026;8(1):100931. doi.org/10.1016/j.landig.2025.100931
16. Sevgi M, Antaki F, Keane PA. Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities. *Br J Ophthalmol*. 2024;108(10):1354-1361. doi.org/10.1136/bjo-2023-325046
17. Rotenstein LS, Holmgren AJ, Thombly R, et al. Changes in clinician time expenditure and visit quantity with adoption of artificial intelligence-powered scribes: A multisite study. *JAMA*. 2026;335(9):1408-1417. doi.org/10.1001/jama.2026.2253
18. Kumar A, Wang H, Muir KW, Mishra V, Engelhard M. A cross-sectional study of GPT-4-based plain language translation of clinical notes to improve patient comprehension of disease course and management. *NEJM AI*. 2025;2(2). doi.org/10.1056/Aloa2400402
19. Introducing ChatGPT Health. Accessed April 9, 2026. openai.com/index/introducing-chatgpt-health
20. Sevgi M, Antaki F, Khan AZ, et al. Complementary human-AI clinical reasoning in ophthalmology. Cornell University. October 28, 2025. arxiv.org/abs/2510.22414
21. Harkar S. What is vibe coding? April 2, 2026. Accessed April 9, 2026. www.ibm.com/think/topics/vibe-coding
22. Gallifant J, Afshar M, Ameen S, et al. The TRIPDD-LLM reporting guideline for studies using large language models. *Nat Med*. 2025;31(1):60-69. doi.org/10.1038/s41591-024-03425-5

FARES ANTAKI, MDCM

- Vitreoretinal Surgeon, AI Researcher, Centre hospitalier de l'Université de Montréal, Montréal, Canada
- fares.antaki@umontreal.ca
- Financial disclosure: Owner (SIMA Surgical Intelligence)