

Comparing Ophthalmologist and Artificial Intelligence Chatbot Responses to Patient Questions

Mostafa Bondok¹, Rishika Selvakumar², Christine Law³, Edsel B Ing^{4,5}, Nupura K Bakshi⁵⁻⁷, Tina Felfeli^{5,8}

¹Section of Ophthalmology, Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary, Canada; ²School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada; ³Department of Ophthalmology, School of Medicine, Queen's University, Kingston, ON, Canada; ⁴Department of Ophthalmology and Visual Sciences, University of Alberta, Edmonton, AB, Canada; ⁵Department of Ophthalmology and Visual Sciences, University of Toronto, Toronto, ON, Canada; ⁶Department of Ophthalmology, St. Michael's Hospital, Unity Health Toronto, Toronto, ON, Canada; ⁷Department of Ophthalmology, Mount Sinai Hospital, Toronto, ON, Canada; ⁸The Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

Correspondence: Tina Felfeli, Department of Ophthalmology and Visual Sciences, University of Toronto, Toronto Western Hospital, 399 Bathurst Street, 6-East Room 432, Toronto, ON, M5T 2S8, Canada, Tel +1 647 678 1634, Fax +1 416 340 3459, Email tina.felfeli@mail.utoronto.ca

Purpose: We evaluated the ability of ChatGPT, an Artificial Intelligence (AI) Chatbot, to respond to patient eye health queries.

Methods: A retrospective, cross-sectional analysis of eye health questions and physician responses posted on the American Academy of Ophthalmology (AAO) "Ask an Ophthalmologist" forum was performed on a random sample from January 2016 to December 2022. We compared board-certified ophthalmologists' responses to ChatGPT (version GPT-4o, OpenAI) responses in September 2024. Primary outcomes included ophthalmologist-rated accuracy of ChatGPT and AAO responses using a 7-point Likert scale, as well as ophthalmologists' preferences between the two responses. Secondary outcomes assessed differences in readability, empathy, and response length between ChatGPT and ophthalmologists.

Results: A random sample 250 questions and responses from 41 board-certified ophthalmologists were evaluated. ChatGPT and AAO responses had similar mean accuracy ratings (5.8 [SD=1.1] vs 5.5 [SD=1.1], $p=0.07$). Evaluators preferred ChatGPT over physician responses in half (49.5%) the cases. Ophthalmologist responses were easier to understand, with a lower mean Flesch-Kincaid Grade Level (Grade 11.0 [SD=2.7] vs Grade 12.7 [SD=1.9], $p<0.001$). Ophthalmologist responses were also significantly shorter than ChatGPT responses (80.6 [SD=56.4] words vs (337.8 [SD=141.6] words, $p<0.001$). Empathy ratings did not differ significantly between ChatGPT and ophthalmologists (4.4 [SD=0.6] vs 4.4 [SD=0.6], $p=0.5$).

Conclusion: Our findings suggest that Chatbot responses were as frequently preferred as physician responses, rated with higher accuracy, and demonstrated comparable empathy in addressing online patient eye health queries. AI chatbots may assist in drafting initial responses to patient concerns, potentially improving efficiency and reducing physician workload.

Keywords: ophthalmology, artificial intelligence, language processing, health education

Introduction

The introduction of innovative artificial intelligence (AI) digital technologies at a rapid rate has provided healthcare workers with a new potential opportunity for more efficient and comprehensive care for patients.¹ Innovative ways to utilize telehealth, AI, and machine learning in the field of ophthalmology have shown considerable promise and capacity to improve health access.¹ Concomitantly, the transition to virtual care during the COVID-19 pandemic has led to an increase in clinicians' time spent on electronic health records (EHR) addressing patient messages,² which may increase physician burnout.

Patients often turn to online resources for health information. In the United States, approximately 80% of internet users rely on the web for health information.³ The emergence of large language models (LLM), such as ChatGPT,⁴ may serve to corroborate online ocular health resources and information from eye specialist consultations to help patients find their answers more quickly and efficiently,⁵ while providing empathetic responses.⁶ Currently, ChatGPT boasts over 180 million users.⁷

Emerging studies have illustrated ChatGPT's ability to generate ophthalmic differential diagnoses,⁸ answer patient health questions,⁹ and perform well on formal ophthalmology examinations.^{10,11} Furthermore, studies have demonstrated how AI can even be used to employ automated assessment, such as when screening articles within systematic reviews,^{12,13} or for the screening, diagnosis, and monitoring of ocular pathologies.^{14–18} A comparison between physician and ChatGPT responses to general health questions on Reddit, an online social media forum, found ChatGPT responses to be of higher quality, empathy, and rated more preferably by physicians.⁹ Bernstein et al, analyzed questions and physician responses on The Eye Care Forum, and found that ChatGPT responses did not differ significantly from ophthalmologist-written responses and were difficult to distinguish.¹⁹ Lyons et al, found that ChatGPT-4 was able to diagnose and triage 44 de novo ophthalmology clinical vignettes with comparable accuracy to ophthalmology trainees at a single centre.²⁰

While many online forums exist to ask physicians health-related questions, the American Academy of Ophthalmology (AAO) “Ask an Ophthalmologist” forum is an accredited platform for patients to ask ophthalmologists about their eye health.²¹ In this retrospective cross-sectional study, we assessed the accuracy, similarity, readability, empathy, and length of ChatGPT's (version GPT-4o, OpenAI) responses to patient questions in comparison to ophthalmologists' responses on a public AAO forum.

Materials and Methods

This is a retrospective cross-sectional study of patient questions and ophthalmologist responses from the “Ask an Ophthalmologist” forum from January 2016 to December 2022. In accordance with AAO Terms of Service, the data was anonymized, and formal permission was obtained from the AAO to use the forum data for this study.²² The AAO granted approval and permitted the input of up to 250 questions into ChatGPT for comparison with ophthalmologist responses. We did not report any identifying information, including physician names in this study. Each patient question was inserted directly into a new ChatGPT session without editing the wording, grammar, or spelling of said question (in September 2024). ChatGPT was selected for this study as it was the most widely used publicly accessible large language model among patients at the time, providing the most relevant platform for assessing AI-generated responses to patient questions. Replies which referred patients to other resources or videos to find the answer to their question were excluded. Unanswered questions were excluded as a comparison could not be made. ChatGPT responses were anonymized by removing revealing information (eg, “As an AI language model...”). This study was exempted from requiring ethics approval by the University of Toronto Research Ethics Board (REB) as it utilized publicly available information with no expectation of privacy.

The readability of ophthalmologist and ChatGPT (version GPT-4o, OpenAI) responses to patient questions were assessed using the Flesch-Kincaid Grade Level, Flesch Reading Ease score, and Gunning Fog Index.²³ To compare response accuracy, two board-certified ophthalmologists (EI, CL) independently evaluated 100 patient questions and their anonymized responses from both sources using a 7-point Likert scale. They rated accuracy based on agreement with the statement: “The response provided is accurate” (1=Strongly disagree, 2=Disagree, 3=Somewhat disagree, 4=Neither agree nor disagree, 5=Somewhat agree, 6=Agree, 7=Strongly agree). Evaluators also indicated which response they preferred.

To assess empathy, graders rated responses based on agreement with the statement: “The response provided is empathetic” using the same Likert scale. Characteristics of empathetic responses included acknowledging the user's frustration, confusion, or concern, providing reassurance, providing guidance on all components of a user's health-related query, and demonstrating support for the user.²⁴ Ratings of accuracy, and empathy were subsequently converted to a numerical scale ranging from 1 to 7, and the mean score between the two raters was used.

Text similarity between ophthalmologists and ChatGPT responses were also compared using CopyLeaks, which is an AI-powered tool used to compare the extent of similarity by categorizing text as “identical” (exact word-for-word matches), “minor changes” (minor variations in a sentence but with the same meaning), and paraphrased text (re-written using different words or sentence structures while retaining the same core idea).²⁵

Statistical Analysis

The distribution of continuous variables was examined for normality using a histogram of data spread, Q–Q plots, and the Shapiro–Wilk test. After assessing for assumptions of normality and similar variance, the paired samples *t*-test was

Table 1 Readability of ChatGPT and Ophthalmologists' Responses to Patient Questions

Flesch-Kincaid Grade Level				Flesch Reading Ease Score						Gunning Fog Index				
AAO		ChatGPT		P value*	AAO		ChatGPT		P value*	AAO		ChatGPT		P value*
Mean	SD	Mean	SD		Mean	SD	Mean	SD		Mean	SD	Mean	SD	
11	2.7	12.7	1.9	<0.001	50.2	14.0	38.2	10.4	<0.001	14.7	3.9	15.2	2.3	<0.001

Notes: *P values are from the paired samples t-test.

Abbreviations: AAO, American Academy of Ophthalmology; SD, Standard Deviation.

used to compare the readability, empathy, accuracy, and number of words in ophthalmologists and ChatGPT responses.²⁶ If these assumptions were not met, the Wilcoxon signed-rank test was used instead. Statistical analyses were conducted using R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria). All tests were two-tailed, and P values less than 0.05 were considered statistically significant.

Results

A total of 1079 questions and responses from 41 ophthalmologists within 30 designated subtopics over the study (2016–2022) period were considered, after 2 questions were excluded. One question was excluded as no response was provided, and the other because the patient was provided a link to an existing video to find the answer to their question. All physicians were board certified ophthalmologists with either an MD (40/41, 97.6%) or DO (1/41, 2.4%), with an average of 28.7 years in practice. Subspecialty representation included retina/vitreoretinal surgery (9/41), comprehensive ophthalmology (8/41), cornea (7/41), glaucoma (6/41), pediatric ophthalmology and strabismus (6/41), and oculoplastics (4/41). The average length of patient questions was 34.1 words (SD=25.1). The number of responses from any single author ranged from answering between 1 and 151 patient questions. Most responses were tagged with more than one subtopic (983/1079, 91.1%). The most tagged topics were “Surgery” (296/1519, 19.5%), “Cataracts” (219/1519, 14.4%), “Glasses, Contacts and Vision Correction” (126/1519, 8.3%), and “General Eye Health” (125/1519, 8.2%). The mean number of words in ophthalmologist responses were significantly lower than ChatGPT responses (80.6 [SD=56.4] words vs 337.8 [SD=141.6] words, $p<0.001$).

The results indicated that AAO responses were easier to read and could be understood by patients with a lower level of education. AAO responses had a higher mean Flesch Reading Ease Score than ChatGPT responses (50.2 [SD=14.0] vs 38.2 [SD=10.4], $p<0.001$) and a lower mean Flesch-Kincaid Grade Level (Grade 11.0 [SD=2.7] vs Grade 12.7 [SD=1.9], $p<0.001$, [Table 1](#)), making them more accessible. Similarly, the Gunning Fog Index showed that AAO responses required a lower reading level compared to ChatGPT responses (Grade 14.7 [SD=3.9] vs Grade 15.2 [SD=2.3], $p<0.001$, [Table 1](#)).

The mean text similarity between ophthalmologist and ChatGPT responses, as measured by CopyLeaks, was less than 1%. Only three responses exhibited any similarity (mean=33.17%), all of which were classified as “paraphrased” rather than “identical” or containing only “minor changes”.

The mean accuracy of ChatGPT responses was comparable to AAO responses (5.8 [SD=1.1] vs 5.5 [SD=1.1], $V=1087.5$, $p=0.07$; [Figure 1](#)). Evaluators preferred ChatGPT over physician responses in approximately half (49.5%) of cases. Empathy ratings did not differ significantly between ChatGPT and AAO responses (4.42 [SD=0.57] vs 4.38 [SD=0.58], $V=1265.5$, $p=0.5$), as shown in [Figure 2](#).

Discussion

In this study, ChatGPT demonstrated slightly higher accuracy but was preferred at a similar rate to ophthalmologists' responses. However, response preference did not always align with accuracy, as more detailed, textbook-like responses were sometimes rated as less appropriate for patients. In some cases, responses with lower accuracy ratings were preferred because they were clearer and more suitable for patients. Similarly, Nanji et al, compared ChatGPT to other online materials for providing postoperative patient instructions, and found that while ChatGPT provided comparable

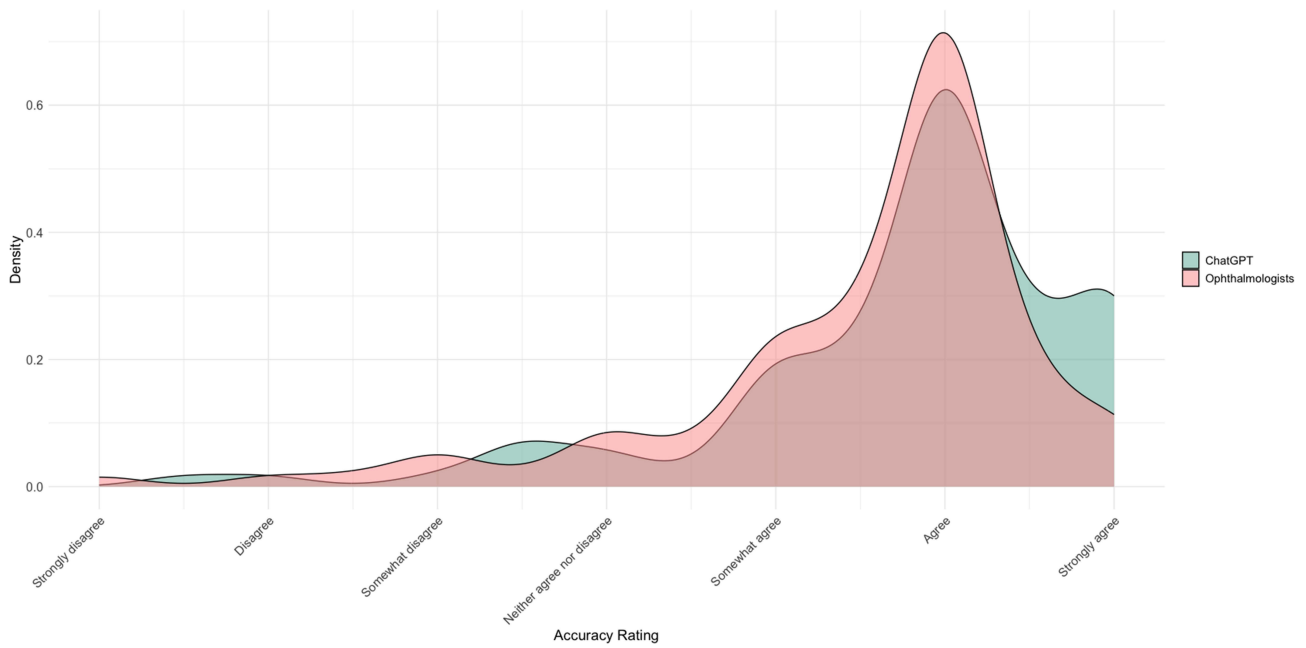


Figure 1 Distribution of mean accuracy ratings for ophthalmologists and ChatGPT responses.

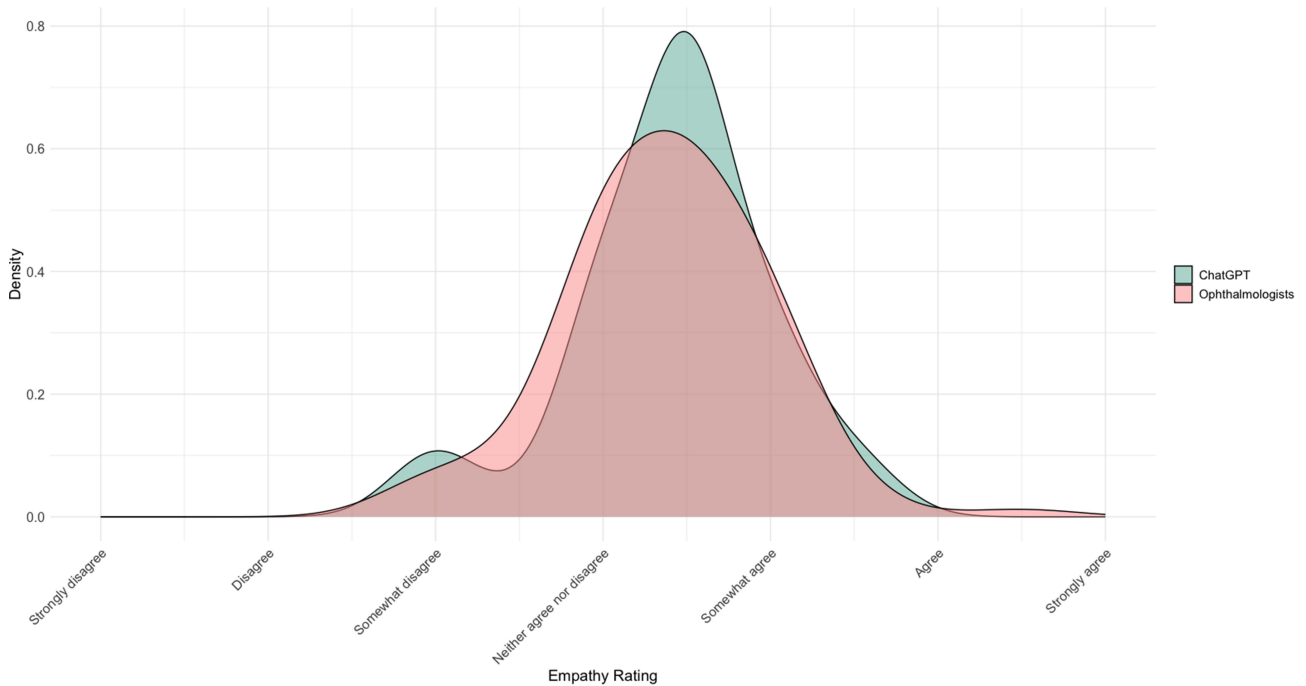


Figure 2 Distribution of mean empathy ratings for ophthalmologists and ChatGPT responses.

procedure-specific information, its responses were less understandable.²⁷ Our analysis showed that physician responses on the AAO “Ask an Ophthalmologist” forum were generally easier to comprehend, required a lower reading grade level, and were shorter than ChatGPT responses. Notably, users can prompt ChatGPT to simplify its language (eg, “please use simpler language”) if a response is too complex. This adaptability has been successfully applied in other healthcare settings, such as simplifying radiological reports for patient education.²⁸

Previous studies have shown that physicians have difficulty differentiating between chatbot and human-written content,^{12,29–32} including responses to online patient questions.¹⁹ This high degree of similarity suggests that ChatGPT responses could serve as templates, allowing physicians to make minor edits before sending them to patients. Although not formally assessed, evaluators in this study also reported recognizing ChatGPT responses based on writing style and response length. In our study, no instances of “hallucinations” or fabricated information were observed. However, prior literature has documented that large language models, including ChatGPT, can generate inaccurate or misleading information, particularly when addressing complex or specialized medical questions.^{33–35} This phenomenon likely stems from the models’ reliance on pattern recognition in text rather than true domain-specific understanding. Consequently, chatbots should not be used to answer patient questions without oversight, and physicians should be aware of the potential medico-legal implications of disseminating unvetted or incorrect information.^{36,37}

The use of online chatbots or AI technology has been widely implemented in other industries to reduce workload burden and improve efficiency,^{38,39} and similar strategies can be implemented in medicine. In the context of health information, the consequences of inaccurate information carry greater risks.^{10,40} Concerns around information accuracy on ChatGPT are valid, but one must consider that alternatives to ChatGPT from the patient perspective include finding this information online. Studies evaluating ocular information online have raised similar concerns about quality and accuracy.⁴¹ Within our dataset, we noted that after providing medical advice, ChatGPT generally recommended that users bring up their concerns to their physicians or eye specialists for further clarification or investigation of a complaint when warranted. Thus, ChatGPT usage by patients serves to corroborate care delivered by physicians, as patients find the abundance of health information on the web overwhelming, conflicting, and confusing.⁴² ChatGPT may help make health information more easily accessible to patients.

Applications of AI in ophthalmology, and in particular deep learning (DL),^{15,43,44} has shown tremendous potential in the screening, diagnosis, and monitoring of ocular disease progression.^{14–18,43,44} These include AI-based detection of retinal fluid in spectral domain OCT,¹⁸ AI-enabled monitoring glaucoma disease progression and severity,¹⁷ and detection of diabetic retinopathy.¹⁶ Similar to imaging-based machine learning applications, when using AI-based responses to patient questions, black-box limitations apply, as it is unclear to what extent ChatGPT is able to evaluate and prioritize sources of information. In our study, the use of ophthalmologist-rated accuracy may also present additional biases.

Our findings are consistent with other studies on the applications of ChatGPT in patient medical education.^{8–11} Ayers et al, evaluated physician responses to general patient health questions, and found ChatGPT responses to patient questions to be shorter and more empathetic,⁹ while our study found ChatGPT responses to be longer and similar in empathy. This may be due to differences in how empathy was judged by raters, differences in comparison groups, and other methodological differences. For instance, Yılmaz et al employed a sentiment-analysis approach, categorizing the emotional tone and attitude of responses using automated and manual approaches, rather than using structured empathy scoring.⁴⁵ Their results similarly highlight that AI-generated ChatGPT responses conveyed more supportive and instructive emotional content. Responses in our study were extracted from the AAO forum, while Ayers et al, utilized responses on a social media forum called Reddit.⁹ Bernstein et al, compared ChatGPT and ophthalmologist responses to patient questions, and found human-rated response accuracy did not differ significantly, and physicians had a difficult time distinguishing human and AI responses.¹⁹ Several other studies have demonstrated the difficulty of differentiating human and AI responses.^{12,29,30}

Previous studies have demonstrated that chatbot performance may also vary across ophthalmic subspecialties.^{45–50} For example, a comparison of ChatGPT, Bard, and Copilot against a trusted patient-information resource (AAPOS) found that, although chatbots demonstrated potential, the AAPOS website consistently outperformed them in both accuracy and readability.⁴⁵ Additionally, chatbot performance has been shown to differ across platforms, highlighting variability in responses depending on the AI model and interface used.^{20,47,49}

One must consider the limitations and biases of these tools when used in isolation, such as the provision of inaccurate information.⁴⁰ Accordingly, the potential benefit of chatbot usage by physicians may be limited to drafting responses to patient questions that the physician can then review, thus reduce physician workload.⁹ While the utility of generative AI models in ophthalmology were initially criticized due to their inability to process images,¹⁰ and for being trained on a dataset using information up until 2021,⁵¹ newer generative AI models, including ChatGPT-4o, are capable of image processing and have

access to the most contemporary information on a topic.⁵² The medico-legal and ethical implications of using ChatGPT in patient communication must also be considered. ChatGPT may generate inappropriate or inaccurate medical advice,³⁶ and, lacking legal personhood, the responsibility for any resulting decisions ultimately rests with the user.³⁷

Limitations and Future Directions

This study compared chatbot to ophthalmologist responses from a single online forum, thus limiting the generalizability of the findings. It is also likely that the publicly available online forum used in this study was incorporated into the training dataset for ChatGPT. In addition, it is unclear exactly how ChatGPT processes information and determines which sources of information are credible. This study evaluated only ChatGPT, which, while being one of the most widely used and publicly accessible large language models at the time of analysis, represents just one of several available platforms. As such, the findings may not be generalizable to other models. Furthermore, all responses were evaluated by board-certified ophthalmologists rather than by patients themselves. While expert assessment allows for objective evaluation of accuracy, readability, and empathy, it may not fully capture patient perspectives or experiences, potentially limiting the interpretation of real-world utility. Future studies should compare the time ophthalmologists spend responding to patients' questions with or without prior ChatGPT-generated responses, to quantify the degree to which using these tools may affect physician workload. Future studies involving patients may consider how these new technologies offer utility within various demographics, such as older adults. In addition, the different ways AI-assisted responses can be safely integrated into hospital or clinical settings should be further investigated.

Conclusion

This retrospective analysis demonstrates the feasibility of utilizing AI chatbots to address patient eye health queries. Our findings suggest that chatbots have the potential to reduce physician workload by drafting initial responses to patient ocular concerns and increase efficiency.

Acknowledgment

The abstract of this paper was presented at the 2024 Canadian Ophthalmological Society (COS) Annual Meeting with interim findings. The abstract was published in the COS Practice Resource Centre: <https://www.cosprc.ca/wp-content/uploads/2024/06/COS-2024-Paper-Abstracts-1.pdf>. Permission was obtained from the American Academy of Ophthalmology to use the Academy's content.

Funding

This study was supported by the generous funds granted to Dr. Tina Felfeli from Fighting Blindness Canada.

Disclosure

The authors have no financial or proprietary interest in any materials discussed in this article.

References

1. Li JP, Liu H, Ting DSJ, et al. Digital technology, tele-medicine and artificial intelligence in ophthalmology: a global perspective. *Prog Retin Eye Res.* 2021;82:100900. doi:10.1016/j.preteyeres.2020.100900
2. Holmgren AJ, Downing NL, Tang M, Sharp C, Longhurst C, Huckman RS. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record use. *J Am Med Inf Assoc.* 2022;29(3):453–460. doi:10.1093/jamia/ocab268
3. Pew Research Center. *Health Topics.* 2011. Available from: <https://www.pewresearch.org/internet/2011/02/01/health-topics-4/>. Accessed May 2, 2023.
4. OpenAI. Introducing ChatGPT. November 30, 2022. Available from: <https://openai.com/blog/chatgpt>. Accessed May 15, 2023.
5. Ting DSJ, Tan TF, Ting DSW. ChatGPT in ophthalmology: the dawn of a new era? *Eye.* 2023;2023:1–4. doi:10.1038/s41433-023-02619-4
6. Graber-Stiehl I. Is the world ready for ChatGPT therapists? *Nature.* 2023;617(7959):22–24. doi:10.1038/D41586-023-01473-4
7. Duarte F. Number of ChatGPT users (Aug 2024). Exploding topics. Available from: <https://explodingtopics.com/blog/chatgpt-users>. Accessed July 2, 2023.
8. Balas M, Ing EB. Conversational AI models for ophthalmic diagnosis: comparison of ChatGPT and the Isabel Pro differential diagnosis generator. *JFO Open Ophthalmol.* 2023;1:100005. doi:10.1016/j.jfop.2023.100005
9. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med.* 2023;183(6):589–596. doi:10.1001/jamainternmed.2023.1838

10. 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. *Ophthalmology Sci.* 2023;3(4):100324. doi:10.1016/J.XOPS.2023.100324
11. Cai LZ, Shaheen A, Jin A, et al. Performance of generative large language models on ophthalmology board style questions. *Am J Ophthalmol.* 2023. doi:10.1016/j.ajo.2023.05.024
12. Mahuli SA, Rai A, Mahuli AV, Kumar A. Application ChatGPT in conducting systematic reviews and meta-analyses. *Br Dent J.* 2023;235(2):90–92. doi:10.1038/S41415-023-6132-Y
13. Blaizot A, Veettil SK, Saidoung P, et al. Using artificial intelligence methods for systematic review in health sciences: a systematic review. *Res Synth Methods.* 2022;13(3):353–362. doi:10.1002/JRSM.1553
14. Grzybowski A, Brona P, Lim G, et al. Artificial intelligence for diabetic retinopathy screening: a review. *Eye.* 2020;34(3):451–460. doi:10.1038/s41433-019-0566-0
15. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol.* 2019;103(2):167–175. doi:10.1136/BJOPHTHALMOL-2018-313173
16. Lim JI, Regillo CD, Sada SVR, et al. Artificial intelligence detection of diabetic retinopathy: subgroup comparison of the eyeart system with ophthalmologists' dilated examinations. *Ophthalmology Sci.* 2023;3(1):100228. doi:10.1016/j.xops.2022.100228
17. Yousefi S, Elze T, Pasquale LR, et al. Monitoring glaucomatous functional loss using an artificial intelligence-enabled dashboard. *Ophthalmology.* 2020;127(9):1170–1178. doi:10.1016/j.ophtha.2020.03.008
18. Keenan TDL, Clemons TE, Domalpally A, et al. Retinal specialist versus artificial intelligence detection of retinal fluid from OCT: age-related eye disease study 2: 10-year follow-on study. *Ophthalmology.* 2021;128(1):100–109. doi:10.1016/j.ophtha.2020.06.038
19. Bernstein IA, Zhang YV, Govil D, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Network Open.* 2023;6(8):e2330320. doi:10.1001/JAMANETWORKOPEN.2023.30320
20. Lyons RJ, Arepalli SR, Fromal O, Choi JD, Jain N. Artificial intelligence chatbot performance in triage of ophthalmic conditions. *Can J Ophthalmol.* 2023;59(4):e301–e308. doi:10.1016/J.CJJO.2023.07.016
21. American Academy of Ophthalmology. Ask an Ophthalmologist. Available from: <https://www.aao.org/eye-health/ask-ophthalmologist>. Accessed May 2, 2023.
22. American Academy of Ophthalmology. Terms of Service. Available from: <https://www.aao.org/terms-of-service>. Accessed May 2, 2023.
23. Shah R, Mahajan J, Oydanich M, Khouri AS. A comprehensive evaluation of the quality, readability, and technical quality of online information on glaucoma. *Ophthalmol Glaucoma.* 2023;6(1):93–99. doi:10.1016/J.OGLA.2022.07.007
24. October TW, Dizon ZB, Arnold RM, Rosenberg AR. Characteristics of physician empathetic statements during pediatric intensive care conferences with family members: a qualitative study. *JAMA Network Open.* 2018;1(3):e180351. doi:10.1001/JAMANETWORKOPEN.2018.0351
25. Copyleaks. Text compare. copyleaks. Available from: <https://app.copyleaks.com/text-compare>. Accessed September 25, 2024.
26. Schober P, Bossers SM, Schwarte LA. Special article: statistical significance versus clinical importance of observed effect sizes: what do P values and confidence intervals really represent? *Anesth Analg.* 2018;126(3):1072. doi:10.1213/ANE.0000000000002798
27. Nanji K, Yu CW, Wong TY, et al. Evaluation of postoperative ophthalmology patient instructions from ChatGPT and google search. *Can J Ophthalmol.* 2023;59(1):e69–e71. doi:10.1016/J.CJJO.2023.10.001
28. Jeblick K, Schachtner B, Dexl J, et al. ChatGPT makes medicine easy to swallow: an exploratory case study on simplified radiology reports. *ArXiv.* 2022. doi:10.48550/arXiv.2212.14882
29. Anderson N, Belavy DL, Perle SM, et al. AI did not write this manuscript, or did it? Can we trick the AI text detector into generated texts? The potential future of ChatGPT and AI in sports & exercise medicine manuscript generation. *BMJ Open Sport Exerc Med.* 2023;9:e001568. doi:10.1136/BMJSEM-2023-001568
30. Dunn C, Hunter J, Steffes W, et al. Artificial intelligence-derived dermatology case reports are indistinguishable from those written by humans: a single-blinded observer study. *J Am Acad Dermatol.* 2023;89(2):388–390. doi:10.1016/j.jaad.2023.04.005
31. Else H. Abstracts written by ChatGPT fool scientists. *Nature.* 2023;613(7944):423. doi:10.1038/D41586-023-00056-7
32. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv.* 2022. doi:10.1101/2022.12.23.521610
33. Goddard J. Hallucinations in ChatGPT: a cautionary tale for biomedical researchers. *Am J Med.* 2023;136(11):1059–1060. doi:10.1016/j.amjmed.2023.06.012
34. 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
35. Kumar M, Mani UA, Tripathi P, Saalim M, Roy S. Artificial hallucinations by google bard: think before you leap. *Cureus.* 2023;15(8):e43313. doi:10.7759/CUREUS.43313
36. Wang C, Liu S, Yang H, Guo J, Wu Y, Liu J. Ethical considerations of using ChatGPT in health care. *J Med Internet Res.* 2023;25(25):e48009. doi:10.2196/48009
37. Zhang J, Zhang Z. Ethics and governance of trustworthy medical artificial intelligence. *BMC Med Inform Decis Mak.* 2023;23(1). doi:10.1186/S12911-023-02103-9
38. Ranoliya BR, Raghawanshi N, Singh S. Chatbot for university related FAQs. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI).* 2017:1525–1530. doi:10.1109/ICACCI.2017.8126057.
39. Brandtzaeg PB, Følstad A. Chatbots: changing user needs and motivations. *Interactions.* 2018;25(5):38–43. doi:10.1145/3236669
40. Nath S, Marie A, Ellershaw S, Korot E, Keane PA. New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. *Br J Ophthalmol.* 2022;106(7):889–892. doi:10.1136/BJOPHTHALMOL-2022-321141
41. Park S, Moskowitz C, Moon JY, Geddie B, Walsh E, Rosenberg JB. Accuracy of online health information on amblyopia and strabismus. *J AAPOS.* 2019;23(6):341–344. doi:10.1016/J.JAPOS.2019.09.007
42. McMullan M. Patients using the Internet to obtain health information: how this affects the patient–health professional relationship. *Patient Educ Couns.* 2006;63(1–2):24–28. doi:10.1016/J.PEC.2005.10.006
43. Ahuja AS, Wagner I, Dorairaj V, Checo S, Hulzen L, Ten R. Artificial intelligence in ophthalmology: a multidisciplinary approach. *Integr Med Res.* 2022;11(4):100888. doi:10.1016/J.IMR.2022.100888

44. Du XL, Li WB, Hu BJ. Application of artificial intelligence in ophthalmology. *Int J Ophthalmol.* 2018;11(9):1555–1561. doi:10.18240/IJO.2018.09.21
45. Yılmaz İE, Berhuni M, Özer Özcan Z, Doğan L. Chatbots talk strabismus: can AI become the new patient educator? *Int J Med Inform.* 2024;191:105592. doi:10.1016/J.IJMEDINF.2024.105592
46. Caranfa JT, Bommakanti NK, Young BK, Zhao PY. Accuracy of vitreoretinal disease information from an artificial intelligence chatbot. *JAMA Ophthalmol.* 2023;141(9):906–907. doi:10.1001/JAMAOPHTHALMOL.2023.3314
47. Cheong KX, Zhang C, Tan TE, et al. Comparing generative and retrieval-based chatbots in answering patient questions regarding age-related macular degeneration and diabetic retinopathy. *Br J Ophthalmol.* 2024;108(10):1443–1449. doi:10.1136/BJO-2023-324533
48. Maywood MJ, Parikh R, Deobhakta A, Begaj T. Performance assessment of an artificial intelligence chatbot in clinical vitreoretinal scenarios. *Retina.* 2024;44(6):954–964. doi:10.1097/IAE.0000000000004053
49. Doğan L, Yılmaz İE. The performance of ChatGPT-4 and Bing chat in frequently asked questions about glaucoma. *Eur J Ophthalmol.* 2025;35(4):1323–1328. doi:10.1177/11206721251321197
50. Özer Özcan Z, Doğan L, Yılmaz İE. Artificial doctors: performance of chatbots as a tool for patient education on keratoconus. *Eye Contact Lens.* 2025;51(3):e112–e116. doi:10.1097/ICL.0000000000001160
51. Lecler A, Duron L, Soyer P. Revolutionizing radiology with GPT-based models: current applications, future possibilities and limitations of ChatGPT. *Diagn Interv Imaging.* 2023;104(6):269–274. doi:10.1016/J.DIII.2023.02.003
52. OpenAI. GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. 2023. Available From: <https://openai.com/gpt-4>. Accessed July 1, 2023.

Clinical Ophthalmology

Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-ophthalmology-journal>

Dovepress

Taylor & Francis Group