# Artificial Intelligence Chatbot Use in Ophthalmology

Riley J Lyons and Sruthi R Arepalli

*Department of Ophthalmology, Emory University School of Medicine, Atlanta, GA, USA*

Introduction: Recent advancements in publicly available artificial intelligence (AI) large language models (LLMs) have captured the interest of the general population and medical community. The ability for large language models to generate natural language text in response to free-text inputs is appealing for multiple reasons, and AI chatbots using this technology have many potential medical and scientific applications. Multiple conversational agents (chatbots) exist, such as Chat Generative Pre-Trained Transformer (ChatGPT; OpenAI), Microsoft's Bing AI (Microsoft Corporation), and Google Bard (Google LLC). As AI chatbot usage increases amongst both laypeople and physicians, it is important to understand their applications within the medical community. Method: A review of the literature was carried out from July 2023 to September 2023 via the PubMed electronic database. The search was restricted to English-language, full-text, peer-reviewed, articles reporting original work on the use of AI chatbots in ophthalmology. The search strategy used was: "("artificial intelligence" OR "AI") AND ("chatbot" OR "GPT" OR "ChatGPT" OR "Bing" OR "Bard" OR "language model") AND ("ophthalmology" OR "ophthalmic" OR "eye" OR "eyecare" OR "ophthalmologist" OR "cornea" OR "uveitis" OR "retina" OR "glaucoma")." Results: The query identified 51 articles; after inclusion and exclusion criteria were applied, 21 articles were yielded. Conclusions: Three potential ophthalmic applications of AI chatbots investigated in the currently available literature include their ability to answer fact-based questions, act as clinical tools, and assist in the production of research. Although several studies show that updated versions of AI chatbots continue to improve their accuracy when answering ophthalmology-related questions, there are still several areas for improvement. In order to improve reliability, AI chatbots must continue to improve in their accuracy, develop the ability to provide links to primary literature and illustrate an ability to capture newer scientific discoveries.

**Keywords**

Artificial intelligence, chatbot, large language models, language models, ophthalmology

**Disclosures:** Riley Lyons and Sruthi Arepalli have no financial or non- financial relationships or activities to declare in relation to this article.

**Review Process:** Double-blind peer review

**Compliance with ethics:** This article involves a review of the literature and did not involve any studies with human or animal subjects performed by any of the authors.

**Data availability:** Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study or during the writing of this article.

**Authorship:** All named authors meet the criteria of the International Committee of Medical Journal Editors for authorship for this manuscript, take responsibility for the integrity of the work as a whole and have given final approval for the version to be published.



**Received:** 6 September 2023

**Accepted:** 2 November 2023

**Published online:** 13 February 2024

**Citation:** *touchREVIEWS in Ophthalmology*. 2024;18(1):9-13

**Corresponding author:** Sruthi R Arepalli, Emory Eye Center, Emory University, 1365B Clifton Road, NE, Suite 2400, Atlanta, GA 30307, USA.
E: sruthiarepalli@gmail.com

**Support:** No funding was received in the publication of this article.

Recent advancements in publicly available artificial intelligence (AI) large language models (LLMs) have piqued the interest of the medical community. LLMs are 'trained' using massive datasets and can generate natural language text in response to free-text inputs. Conversational agents (chatbots), such as Chat Generative Pre-Trained Transformer (ChatGPT; OpenAI), Microsoft's Bing AI (Microsoft Corporation), and Google Bard (Google LLC), utilize these LLMs. These AI chatbots are easy to use and can produce complex responses to free-text prompts that are often not easily distinguishable from those of humans.

Physicians across all medical fields have demonstrated interest in AI chatbots' ability to perform tasks that typically require significant time investment from knowledgeable, trained humans. Indeed, many ophthalmologists have already published manuscripts detailing their experiences testing novel applications of various AI chatbots. Physicians have displayed both excitement and skepticism regarding the potential scientific and medical applications of these AI chatbots. The purpose of this narrative review is to provide a synthesis of published studies evaluating the utility of AI chatbots in ophthalmology.

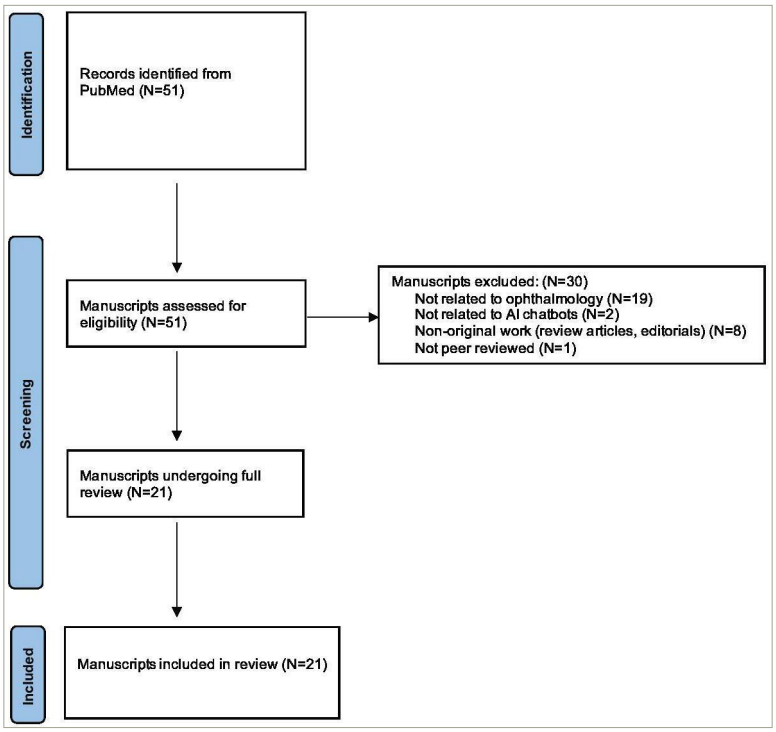
## Method

A review of the literature was carried out from July 2023 to September 2023 via the PubMed electronic database. The publication date cutoff for manuscripts included in our review was September 1, 2023. The search was restricted to English-language, full-text, peer-reviewed, articles reporting original work on the use of AI chatbots in ophthalmology. The search strategy and terms used were: "("artificial intelligence" OR "AI") AND ("chatbot" OR "GPT" OR "ChatGPT" OR "Bing" OR "Bard" OR "language model") AND ("ophthalmology" OR "ophthalmic" OR "eye" OR "eyecare" OR "ophthalmologist" OR "cornea" OR "uveitis" OR "retina" OR "glaucoma")." This study involves a review of the literature and did not involve any studies with human or animal subjects performed by any of the authors. The aims and scope of this strategy were identified by the authors and it is not designed to be a comprehensive summary.

### Study selection

The two named co-authors reviewed titles and abstracts of English, full-text articles for the following inclusion criteria: (1) the study addressed AI chatbot use in ophthalmology and (2) the study reported original, peer-reviewed, research (for example, editorial, correspondence, and review articles were not included). No other exclusion criteria were used for screening aside from those listed above.

Figure 1: Study selection process




*AI = artificial intelligence.*

## Results

The query identified 51 articles. Inclusion and exclusion criteria were applied, eliminating 19 articles not related to ophthalmology, 8 review articles and editorials, 2 articles not related to AI chatbots, and 1 article that was not peer-reviewed. Twenty-one articles remained and underwent full text review (Figure 1).

Among the 21 articles, there were three main functionalities for which the AI chatbots were tested. The most common study type (n=14 studies) evaluated AI chatbots' ability to accurately answer questions related to a variety of topics within ophthalmology (i.e. 'knowledge assessments'). The second most common study type evaluated AI chatbots' ability to function as clinical tools (n=5 studies). The third study type evaluated AI chatbots' ability to function as ophthalmic research tools (n=2 studies).

### AI chatbot knowledge assessment

Fourteen studies in our review performed knowledge assessments of AI chatbots (*Table 1*). While some studies tested chatbots' ability to answer questions on specific ophthalmology diagnoses or subjects, others presented a wide range of questions covering many ophthalmic subspecialties.

Potapenko et al., Rasmussen et al., Caranfa et al., Ali, Biswas et al. and Momenaei et al. presented ChatGPT with questions related to common retinal diseases, vernal keratoconjunctivitis, vitreoretinal disease, lacrimal drainage disorders, myopia and surgical treatment of retinal diseases, respectively.[1–5,7]

Potapenko et al. and Rasmussen et al. found that ChatGPT performed well when presented with questions regarding general disease characteristics; however, in both studies, ChatGPT's performance dropped when asked about management options. The authors found that ChatGPT was consistent with its answers when prompted multiple times with the same question.[1,2] Caranfa et al. queried ChatGPT with 52 questions frequently asked by patients related to vitreoretinal conditions and procedures and found that only 8 (15.4%) questions were graded as "completely accurate". Additionally, when ChatGPT was presented with the same questions 14 days later, the authors found that 26 (50%) responses significantly changed.[3] Potapenko et al., Rasmussen et al., and Caranfa et al. did not document the version of ChatGPT used in their studies.[1–3]

Ali found that ChatGPT (GPT-3.5) answered only 40% of prompts related to lacrimal drainage disorders correctly.[4] Biswas et al. generated 11 questions related to myopia and presented them to ChatGPT (GPT-3.5) five times for grading. The graders rated 48.7% of responses as "good", 24% of responses as "very good," and 5.4% of responses as either "inaccurate" or "flawed."[5] Manuscripts from Potapenko et al., Rasmussen et al., Caranfa et al. and Ali all reported instances of potentially harmful responses by the chatbots.[1–4]

Lim et al. curated 31 myopia-related patient questions and presented them to Google Bard and ChatGPT using GPT-3.5 and GPT-4.0. AI chatbot responses were graded by three experienced ophthalmologists for accuracy and comprehensiveness. Accuracy was graded as either "poor," "borderline," or "good" and comprehensiveness was assessed using a five-point scale from 1–5. AI chatbot responses received grades of "good' in 80.6%, 61.3%, and 54.8% of cases for ChatGPT (GPT-4.0), ChatGPT (GPT-3.5), and Google Bard, respectively, while all chatbots showed high average "comprehensiveness scores" of 4.23, 4.11, and 4.35, correspondingly.[6]

Momenaei et al. found that ChatGPT (GPT-4.0) answered most questions related to vitreoretinal surgeries for retinal detachments (RDs), macular holes (MHs) and epiretinal membranes (ERMs) correctly. Answers were graded as inappropriate for 5.1%, 8.0%, and 8.3% of questions related to

## Table 1: Article summary

| Number | Author | Article title | Chatbot application | AI chatbot type | ChatGPT version |
|---|---|---|---|---|---|
| 1 | Potapenko et al.[1] | Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. | Knowledge assessment | ChatGPT | Not specified |
| 2 | Rasmussen et al.[2] | Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis. | Knowledge assessment | ChatGPT | Not specified |
| 3 | Caranfa et al.[3] | Accuracy of Vitreoretinal Disease Information From an Artificial Intelligence Chatbot. | Knowledge assessment | ChatGPT | Not specified |
| 4 | Ali[4] | ChatGPT and Lacrimal Drainage Disorders: Performance and Scope of Improvement. | Knowledge assessment | ChatGPT | GPT-3.5 |
| 5 | Biswas et al.[5] | Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia. | Knowledge assessment | ChatGPT | GPT-3.5 |
| 6 | Lim et al.[6] | Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. | Knowledge assessment | ChatGPT and Bard | GPT-3.5, GPT-4.0 |
| 7 | Momenaei et al.[7] | Appropriateness and Readability of ChatGPT-4-Generated Responses for Surgical Treatment of Retinal Diseases. | Knowledge assessment | ChatGPT | GPT-4.0 |
| 8 | Antaki et al.[8] | Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of Its Successes and Shortcomings. | Knowledge assessment | ChatGPT | "January 9 legacy version" and "ChatGPT Plus" |
| 9 | Mihalache et al.[9] | Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment. | Knowledge Assessment | ChatGPT | "January 9 version" and "February 13 version" |
| 10 | Mihalache et al.[10] | Performance of an Upgraded Artificial Intelligence Chatbot for Ophthalmic Knowledge Assessment. | Knowledge assessment | ChatGPT | GPT-4.0 |
| 11 | Cai et al.[11] | Performance of Generative Large Language Models on Ophthalmology Board-Style Questions. | Knowledge assessment | ChatGPT and Bing Chat | GPT-3.5, GPT-4.0 |
| 12 | Teebagy et al.[12] | Improved Performance of ChatGPT-4 on the OKAP Examination: A Comparative Study with ChatGPT-3.5. | Knowledge assessment | ChatGPT | GPT-3.5, GPT-4.0 |
| 13 | Moshirfar et al.[13] | Artificial Intelligence in Ophthalmology: A Comparative Analysis of GPT-3.5, GPT-4, and Human Expertise in Answering StatPearls Questions. | Knowledge assessment | ChatGPT | GPT-3.5, GPT-4.0 |
| 14 | Panthier et al.[14] | Success of ChatGPT, an AI language model, in taking the French language version of the European Board of Ophthalmology examination: A novel approach to medical knowledge assessment. | Knowledge assessment | ChatGPT | GPT-4.0 |
| 15 | Tsui et al.[15] | Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. | Clinical tool | ChatGPT | "February 13 version" |
| 16 | Bernstein et al.[16] | Comparison of Ophthalmologist and Large Language Model Chatbot Responses to Online Patient Eye Care Questions. | Clinical tool | ChatGPT | GPT-3.5 |
| 17 | Lyons et al.[17] | Artificial intelligence chatbot performance in triage of ophthalmic conditions. | Clinical tool | ChatGPT and Bing Chat | GPT-4.0 |
| 18 | Hu et al.[18] | What can GPT-4 do for Diagnosing Rare Eye Diseases? A Pilot Study. | Clinical tool | ChatGPT | GPT-4.0 |
| 19 | Waisberg et al.[19] | GPT-4 and Ophthalmology Operative Notes. | Clinical tool | ChatGPT | GPT-4.0 |
| 20 | Valentín-Bravo et al.[20] | Artificial Intelligence and new language models in Ophthalmology: Complications of the use of silicone oil in vitreoretinal surgery. | Research tool | ChatGPT | Not specified |
| 21 | Hua et al.[21] | Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots. | Research tool | ChatGPT | GPT-3.5, GPT-4.0 |

*ChatGPT = Chat Generative Pre-Trained Transformer; OKAP = Ophthalmic Knowledge Assessment Program.*

RD, MH, and ERM, respectively. The authors evaluated the readability of chatbot responses and found overall low readability scores.[7]

In contrast to the above studies focusing on specific diseases or ophthalmic subspecialties, more studies have been published evaluating ChatGPT's ability to respond to board-review style questions across a wide range of ophthalmic subspecialties. Antaki et al. presented review questions from two common online ophthalmology board-review resources, OphthoQuestions™ and Basic and Clinical Science Course (BCSC) Self-Assessment Program, to two ChatGPT versions ("January 9 legacy version" and the newer "ChatGPT Plus"). The 'legacy' version of ChatGPT answered 55.8% and 42.7% of questions correctly in the BSCS and OphthoQuestions sets, respectively. The ChatGPT Plus version increased accuracy to 59.4% and 49.2% of questions answered correct for the BSCS and OphthoQuestions sets, respectively.[8]

Similarly, Mihalache et al. presented 125 questions from OphthoQuestions to ChatGPT in January 2023 and February 2023. The authors then presented the same questions to an upgraded ChatGPT (GPT-4.0) in March 2023. They found that ChatGPT's accuracy improved with each version, from 46% ('January 9' version) to 58% ('February 13' version), and then to 84% (GPT-4.0).[9,10] Cai et al. compared answers to BCSC questions from ChatGPT using GPT-3.5 and GPT-4.0, Bing Chat using GPT-4.0 technology and human respondents. They found that ChatGPT using GPT-3.5 scored poorly (58.5%), while ChatGPT using GPT-4.0, Bing Chat using GPT-4.0 and human respondents scored similarly (71.6%, 71.2%, and 72.7%, respectively).[11] Teebagy et al. also presented BCSC review questions to ChatGPT using GPT-3.5 and GPT-4.0 and found accuracies of 57% and 81% for GPT-3.5 and GPT-4.0, respectively.[12] Moshirfar et al. published a similar study assessing the accuracy of multiple versions of ChatGPT on ophthalmology board review materials while comparing AI chatbot performance with that of humans. For ChatGPT-4.0, ChatGPT-3.5 and humans, accuracy was reported as 73.2%, 55.5% and 58.3%, respectfully. ChatGPT-4.0 was noted to outperform both humans and ChatGPT-3.5 for all question subcategories except for one.[13] Panthier et al. found very high levels of accuracy when presenting European Board of Ophthalmology examination material to ChatGPT-4.0.[14]

## AI chatbots as clinical tools

Significant interest has been expressed regarding the potential clinical applications of AI chatbots. Five of the reviewed studies evaluated various clinical functions of AI chatbots (*Table 1*).

Tsui et al. created 10 prompts representing common patient complaints and presented them to ChatGPT ('February 13 version'). Responses were graded as either "precise or imprecise" and "suitable or unsuitable." The authors found that 8 out of 10 responses were precise and suitable, while two prompt responses were imprecise and unsuitable.[15]

Bernstein et al. presented 200 eye care questions from an online forum to ChatGPT (GPT-3.5) and compared the AI chatbot's responses to those of physicians affiliated with the American Academy of Ophthalmology. Masked ophthalmologist reviewers were randomly presented with either physician or AI chatbots answers, and the reviewers attempted to discern between the human and AI-generated responses. Additionally, reviewers assessed answers for accuracy and potential and severity of harm. The reviewer accuracy for distinguishing between human responses and AI-generated responses was 61%. Additionally, the reviewers rated the human responses and AI-generated responses similarly when assessing for incorrect information and potential for harm.[16]

Lyons et al. created 44 clinical vignettes representing 24 common ophthalmic complaints. Most diagnoses had two prompt versions, one using 'buzz words' and the other using 'generic or layman' language more likely to be used by patients. The authors presented prompts to ChatGPT (GPT-4.0), Bing Chat using GPT-4.0, WebMD Symptom Checker, and ophthalmology trainees. Responses were graded for diagnostic accuracy and appropriateness of triage urgency. Ophthalmology trainees, ChatGPT, Bing Chat, and WebMD provided the appropriate diagnosis in 42 (95%), 41 (93%), 34 (77%), and 8 (33%) cases, respectively. Triage urgency was graded as appropriate for 38 (86%), 43 (98%), and 37 (84%) responses from ophthalmology trainees, ChatGPT, and Bing Chat, respectively. The authors found no cases of harmful responses from ChatGPT and one such case from Bing Chat. ChatGPT did not provide references, while Bing Chat was able to link directly to source material.[17]

Hu et al. assessed ChatGPT's (GPT-4.0) ability to diagnose rare eye diseases. They chose 10 rare eye diseases from an online collection and presented simulated scenarios to ChatGPT for three different end-user test cases, including "patients," "family physicians" and "junior ophthalmologists". For each test case, more information was presented to ChatGPT to simulate the proposed increasing ophthalmologic knowledge base from patients to family physicians to junior ophthalmologists, and responses were graded by senior ophthalmologists for "suitability" of response ("appropriate" or "inappropriate") and for correctness of diagnoses ("right" or "wrong"). For the test cases simulating patients, family physicians and junior ophthalmologists, the answers produced by ChatGPT were graded as appropriate in seven (70%), ten (100%) and eight (80%) cases, respectively, while diagnoses were graded as correct in zero (0%), five (50%) and nine (90%) cases, respectively.[18]

Waisberg et al. tested ChatGPT's (GPT-4.0) ability to write ophthalmic operative notes by asking it to generate a cataract surgery operative note. The authors noted ChatGPT produced a detailed note and listed key steps of cataract surgery; however, the authors did not apply specific grading criteria for analysis.[19]

## AI chatbots as research tools

The third application of AI chatbots noted in our review was their potential use as ophthalmic research tools (*Table 1*). Valentín-Bravo et al. published a report in which the authors asked ChatGPT about the use of silicone oil in vitreoretinal surgery and requested its response be written as a scientific paper with references. They reported ChatGPT produced a coherent, brief summary of the topic, but its response included significant inaccuracies and was not at the level of an experienced ophthalmologist author. When asked for citations, ChatGPT's response included "hallucinated" articles. The authors did not report the version of ChatGPT used in their experiment.[20]

Hua et al. used two versions of ChatGPT (GPT- 3.5 and GPT-4.0) to generate scientific abstracts for clinical research questions across seven ophthalmic subspecialties and then graded abstracts using DISCERN criteria. The abstracts produced by ChatGPT-3.5 and ChatGPT-4.0 received grades of 36.9 and 38.1, respectively, qualifying them as "average-quality" abstracts. The authors noted hallucination rates of 31% and 29% for articles generated by GPT-3.5 and GPT-4.0, respectively. Additionally, available tools designed to detect AI-generated work were unable to reliably identify the abstracts as AI generated.[21]

## Discussion

Our review identified three primary ophthalmic applications for which AI chatbots have been evaluated in the literature including answering knowledge assessments, use as clinical tools and use as research tools. The simplest of these three functions is a chatbot's ability to accurately respond to fact-based questions in a knowledge assessment. If AI chatbots excel at this task, they could potentially serve as educational aids for laypeople, students and physicians. However, if they do not consistently answer questions correctly, they could be a potential source of misinformation.

Amongst the studies evaluated in this review, several studies demonstrated increasing accuracy when comparing ChatGPT-4.0 to its predecessor ChatGPT-3.5. Also, multiple studies found similar accuracies when comparing responses from ChatGPT-4.0 to those of human respondents.[11,13,22]

Several chatbot shortcomings were identified by authors, including inability to reliably link to source texts and occasional 'hallucinations' in which chatbots answered questions with false information. As future versions of AI chatbots are developed, continued evaluations will be necessary to monitor for improvements. Additionally, it will be interesting to see how AI chatbots incorporate new information into their responses, as ophthalmologic advancements occur alongside AI chatbot improvement.

Two more complex potential applications of AI chatbots within ophthalmology are their ability to assist physicians as clinical tools or in performing research. Before entrusting AI chatbots with these complex functions, physicians must first ensure chatbots can consistently and accurately answer fact-based questions as discussed above.

One potential clinical application of AI chatbots explored in three of the reviewed studies above is AI chatbots' ability to assist in medical triage of patient complaints.[16–18] Lyons et al. demonstrated that ChatGPT (GPT-4.0) performed very well triaging representative ophthalmic clinical vignettes, with high levels of diagnostic accuracy and triage urgency.[17] However, the authors do not recommend patients or physicians turn to AI chatbots for this function. In their current state, AI chatbots are not suitable substitutes for contacting your doctor for medical advice. Further studies are necessary to educate ophthalmologists and laypeople on the performance of readily available AI tools that many patients are likely already using for medical advice, despite being instructed otherwise.

Another potentially controversial use of AI chatbots in ophthalmology is their use as research tools. Leaders in ophthalmology and across many medical and scientific fields have condemned the use of AI as scientific writers.[23] However, studies like those by Hua et al. are still necessary to alert and update the medical and scientific communities regarding the capabilities of currently available AI technology.[21]

## Conclusion

The recent rise of AI chatbots has brought about a major disruption to the medical and scientific community. AI chatbots are incredible tools, and more work must be done to further investigate their potential uses and limitations within ophthalmology. Three potential ophthalmic applications of AI chatbots investigated in the currently available literature include their ability to answer fact-based questions, act as clinical tools and assist in the production of research. Although newer iterations of chatbots such as ChatGPT-4.0 have shown significant improvement in their ability to accurately answer ophthalmology-related questions compared to older versions, AI chatbots must continue to improve their accuracy and develop the ability to provide links to primary literature. Also, it is not-yet known how AI chatbots may incorporate new scientific discoveries into their answers. As they currently exist, AI chatbots should not be used as clinical tools or used in the production of scientific writing. However, further studies are warranted to continue to monitor the improvement of chatbots in these areas, and perhaps future advanced chatbot iterations could be designed specifically for ophthalmic applications after extensive validation and safety testing. ❑


1. Potapenko I, Boberg-Ans LC, Stormly Hansen M, et al. Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. *Acta Ophthalmol*. 2023;101:829–31. DOI: 10.1111/aos.15661.
2. Rasmussen MLR, Larsen AC, Subhi Y, Potapenko I. Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis. *Graefes Arch Clin Exp Ophthalmol*. 2023;261:3041–3. DOI: 10.1007/s00417-023-06078-1.
3. Caranfa JT, Bommakanti NK, Young BK, Zhao PY. Accuracy of vitreoretinal disease information from an artificial intelligence chatbot. *JAMA Ophthalmol*. 2023;141:906–7. DOI: 10.1001/jamaophthalmol.2023.3314.
4. Ali MJ. ChatGPT and Lacrimal drainage disorders: Performance and scope of improvement. *Ophthalmic Plast Reconstr Surg*. 2023;39:221–5. DOI: 10.1097/IOP.0000000000002418.
5. Biswas S, Logan NS, Davies LN, et al. Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia. *Ophthalmic Physiol Opt*. 2023;43:1562–70. DOI: 10.1111/opo.13207.
6. Lim ZW, Pushpanathan K, Yew SME, et al. Benchmarking large language models' performances for myopia care: A comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and google bard. *EBioMedicine*. 2023;95:104770. DOI: 10.1016/j.ebiom.2023.104770.
7. Momenaei B, Wakabayashi T, Shahlaee A, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina*. 2023;7:862–8. DOI: 10.1016/j.oret.2023.05.022.
8. Antaki F, Touma S, Milad D, et al. Evaluating the performance of ChatGPT in ophthalmology: An analysis of its successes and shortcomings. *Ophthalmol Sci*. 2023;3:100324. DOI: 10.1016/j.xops.2023.100324.
9. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023;141:589–97. DOI: 10.1001/jamaophthalmol.2023.1144.
10. Mihalache A, Huang RS, Popovic MM, Muni RH. Performance of an upgraded artificial intelligence chatbot for ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023;141:798–800. DOI: 10.1001/jamaophthalmol.2023.2754.
11. Cai LZ, Alabiad C. Performance of generative large language models on ophthalmology board-style questions. *Am J Ophthalmol*. 2023;254:141–9. DOI: 10.1016/j.ajo.2023.07.030.
12. Teebagy S, Colwell L, Wood E, et al. Improved performance of ChatGPT-4 on the OKAP examination: A comparative study with ChatGPT-3.5. *J Acad Ophthalmol*. 2023;15:e184–7. DOI: 10.1055/s-0043-1774399.
13. Moshirfar M, Altaf AW, Stoakes IM, et al. Artificial intelligence in ophthalmology: A comparative analysis of GPT-3.5, GPT-4, and human expertise in answering StatPearls questions. *Cureus*. 2023;15:e40822. DOI: 10.7759/cureus.40822.
14. Panthier C, Gatinel D. Success of ChatGPT, an AI language model, in taking the French language version of the European Board of Ophthalmology examination: A novel approach to medical knowledge assessment. *J Fr Ophtalmol*. 2023;46:706–11. DOI: 10.1016/j.jfo.2023.05.006.
15. Tsui JC, Wong MB, Kim BJ, et al. Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. *Eye*. 2023;37:3692–3. DOI: 10.1038/s41433-023-02556-2.
16. Bernstein IA, Zhang YV, Govil D, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Netw Open*. 2023;6:e2330320. DOI: 10.1001/jamanetworkopen.2023.30320.
17. Lyons RJ, Arepalli SR, Fromal O, et al. Artificial intelligence chatbot performance in triage of ophthalmic conditions. *Can J Ophthalmol*. 2023;S0008-4182(23)00234-X. DOI: 10.1016/j.jcjo.2023.07.016.
18. Hu X, Ran AR, Nguyen TX, et al. What can GPT-4 do for diagnosing rare eye diseases? A pilot study. *Ophthalmol Ther*. 2023;12:3395–402. DOI: 10.1007/s40123-023-00789-8.
19. Waisberg E, Ong J, Masalkhi M, et al. GPT-4 and ophthalmology operative notes. *Ann Biomed Eng*. 2023;51:2353–5. DOI: 10.1007/s10439-023-03263-5.
20. Valentín-Bravo FJ, Mateos-Álvarez E, Usategui-Martín R, et al. Artificial intelligence and new language models in ophthalmology: Complications of the use of silicone oil in vitreoretinal surgery. *Arch Soc Esp Oftalmol (Engl Ed)*. 2023;98:488–9. DOI: 10.1016/j.oftale.2023.06.010.
21. Hua HU, Kaakour AH, Rachitskaya A, et al. Evaluation and comparison of ophthalmic scientific abstracts and references by current artificial intelligence chatbots. *JAMA Ophthalmol*. 2023;141:819–24. DOI: 10.1001/jamaophthalmol.2023.3119.
22. Balas M, Ing EB. Conversational AI models for ophthalmic diagnosis: Comparison of ChatGPT and the Isabel pro differential diagnosis generator. *JFO Open Ophthalmology*. 2023;1:100005. DOI: 10.1016/j.jfop.2023.100005.
23. Bressler NM. What artificial intelligence chatbots mean for editors, authors, and readers of peer-reviewed ophthalmic literature. *JAMA Ophthalmol*. 2023;141:514. DOI: 10.1001/jamaophthalmol.2023.1370.