### Title

Large Language Models in Ophthalmology: A Review of Publications from Top Ophthalmology Journals




### Journal

Ophthalmology Science, 5(3)

### ISSN

2666-9145



### Authors

Agnihotri, Akshay Prashant

Nagel, Ines Doris

Artiaga, Jose Carlo M

et al.



### Publication Date

2025-05-01

### DOI

10.1016/j.xops.2024.100681

AMERICAN ACADEMY OF OPHTHALMOLOGY®

# Large Language Models in Ophthalmology: A Review of Publications from Top Ophthalmology Journals

*Akshay Prashant Agnihotri, MS, DNB,*[1,2,3,*] *Ines Doris Nagel, MD,*[1,2,4,*] *Jose Carlo M. Artiaga, MD, FICO,*[5,6] *Ma. Carmela B. Guevarra, MD,*[7,8] *George Michael N. Sosuan, MD,*[5] *Fritz Gerald P. Kalaw, MD*[1,2,9]

***Purpose:*** To review and evaluate the current literature on the application and impact of large language models (LLMs) in the field of ophthalmology, focusing on studies published in high-ranking ophthalmology journals.

***Design:*** This is a retrospective review of published articles.

***Participants:*** This study did not involve human participation.

***Methods:*** Articles published in the first quartile (Q1) of ophthalmology journals on Scimago Journal & Country Rank discussing different LLMs up to June 7, 2024, were reviewed, parsed, and analyzed.

***Main Outcome Measures:*** All available articles were parsed and analyzed, which included the article and author characteristics and data regarding the LLM used and its applications, focusing on its use in medical education, clinical assistance, research, and patient education.

***Results:*** There were 35 Q1-ranked journals identified, 19 of which contained articles discussing LLMs, with 101 articles eligible for review. One-third were original investigations (32%; 32/101), with an average of 5.3 authors per article. The United States (50.4%; 51/101) was the most represented country, followed by the United Kingdom (25.7%; 26/101) and Canada (16.8%; 17/101). ChatGPT was the most used LLM among the studies, with different versions discussed and compared. Large language model applications were discussed relevant to their implications in medical education, clinical assistance, research, and patient education.

***Conclusions:*** The numerous publications on the use of LLM in ophthalmology can provide valuable insights for stakeholders and consumers of these applications. Large language models present significant opportunities for advancement in ophthalmology, particularly in team science, education, clinical assistance, and research. Although LLMs show promise, they also show challenges such as performance inconsistencies, bias, and ethical concerns. The study emphasizes the need for ongoing artificial intelligence improvement, ethical guidelines, and multidisciplinary collaboration.

***Financial Disclosure(s):*** The author(s) have no proprietary or commercial interest in any materials discussed in this article. *Ophthalmology Science 2025;5:100681 © 2024 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

*Supplemental material available at www.ophthalmologyscience.org.*

Artificial intelligence (AI) is a general term that implies using computers to model advanced behaviors requiring less human intervention.[1] Earlier studies regarding the use of AI in medicine, particularly in ophthalmology, entailed a subfield of AI called machine learning (ML), which served a pivotal role in detecting and advancing the understanding of retinal pathologies.[2–5] Recent investigations in ophthalmology have delved into the potential use of large language models (LLMs), a type of AI model that uses ML to analyze and generate language. Large language models are natural language processing models that leverage ML techniques and are self-trained on large amounts of text data from articles, books, and other internet-based contents to process, analyze, and generate human language. With the training of text dataset, the models learn how words are used and can interpret and generate text with minimal to no specific fine-tuning.[6] ChatGPT (OpenAI) is an online chatbot that became widely available in November 2022 and gained public interest.[7] Since its release, there has been an exponential rise in medical research regarding the potential use of LLMs.

The numerous publications on the use of LLM in ophthalmology can provide valuable insights for stakeholders and consumers of these applications. Therefore, it is imperative that the evidence provided in these publications is both reliable and derived from reputable sources. To achieve this, editors and members of the editorial board of various journals implement stringent criteria and review processes to ascertain which articles fulfill the standards requisite for publication. Findings about relevant articles discussing LLM from trusted and highly ranked ophthalmology journals have not been explored. Because we are in



https://doi.org/10.1016/j.xops.2024.100681
ISSN 2666-9145/25

the era of a fast-paced and rapidly evolving field of technology, it would be reasonable to evaluate the available evidence in order to understand how the field of ophthalmology stands with LLM use. Furthermore, considering that research in computer vision relevant to ophthalmology is presently accessible and will remain so in the foreseeable future, it is advisable to present all the evidence obtained from these esteemed ophthalmology journals. Hence, the objectives of the study are twofold: (1) to evaluate and characterize different LLM-related literature from highly ranked journals in terms of research productivity and (2) to enumerate and summarize the findings of the different literature regarding the implications of the use of LLMs.

## Methods

This study entailed reviewing published literature from ophthalmology-related journals. It did not involve human participation and, hence, did not require institutional review board approval.

### Identification of Ophthalmology Journals and Obtaining Articles of Interest

Ophthalmology journals ranked as first quartile (Q1) in the ophthalmology category from the Scimago Journal & Country Rank (SJR) using the Scimago website platform[8] were identified. Briefly, SJR Q1 journals are journals that are in the top 25% of the category (in our case, ophthalmology). The ranking is based on weighted citations per article in a given journal over a period of 3 years.

After the journals were identified, each journal website platform was examined for available topics of interest. The following topics were searched on June 7, 2024: "Large language model(s)," "ChatGPT," and "Chatbot(s)." All available articles, including accepted articles available as preproof, were included in the study. The University of California, San Diego, Library provided access to full-text articles when available. Full-text articles that require membership or premium subscription or are locked behind a paywall were excluded from the study.

### Literature Review and Analysis of Articles

The articles were examined according to article type, publication date, number of authors, and countries of affiliations of listed authors. Concerning the topic of LLMs, each article was carefully reviewed by all authors and was parsed according to the following data elements when available: purpose of the study, LLM(s) (including versions, if mentioned) used or discussed, application of the LLM, results and findings, and implications of LLM use. The application of LLM was divided into 4 categories: (1) medical education, (2) clinical assistance, (3) research, and (4) patient education. All data elements were tabulated and analyzed using Microsoft Excel v16.58 (Microsoft Corporation), and all figures were generated using Microsoft Excel and Microsoft PowerPoint v16.58 (Microsoft Corporation).

## Results

Thirty-five (35/141) journals from the SJR platform were identified as Q1-ranked journals, 19 of which have publications discussing LLMs. From these 19 journals, 106 articles were retrieved, with 5 articles requiring member or premium subscriptions or were behind a paywall and, hence, were excluded. Thus, a total of 101 articles were reviewed, parsed, and analyzed (Fig 1).

Table 1 presents all articles from 19 journals included in the study. The earliest article discussing LLM in ophthalmology was published in July 2022, and the latest was in June 2024, with noted gradual increase in publication trends (Fig 2). There were various article types published, the most common being original investigations (32/101; 32%) (Table S2, available at https://www.ophthalmologyscience.org). Interestingly, 34% (34/101) of the articles were correspondence, comments, letters, and responses from the articles. There were a total of 528 authors listed from all articles, averaging 5.3 authors per article (Table S3, available at https://www.ophthalmologyscience.org). In terms of countries of affiliation of listed authors, half of the articles (50.4%, 51/101) have contributions from authors affiliated with institutions from the United States, followed by the United Kingdom (25.7%; 26/101) and Canada (16.8%; 17/101). (Table S4, available at https://www.ophthalmologyscience.org).

Review of articles regarding the discussion of LLMs is seen in Table 5. There were 34 identified LLMs discussed from the 101 articles, with ChatGPT versions providing the majority of the LLMs used either for discussion or comparison (ChatGPT-4 [54], ChatGPT-3.5 [40], ChatGPT-3 [7], ChatGPT [12]). In terms of LLM applications, 31 articles discussed the use of LLM for medical education, 32 discussed its use for clinical assistance, 24 discussed its use for research, and 23 discussed its use for patient education. An in-depth review of articles regarding the application of LLM and its implications is further discussed in the discussion section.

## Discussion

The present study revealed several interesting findings. Even though investigations regarding the use of LLM in

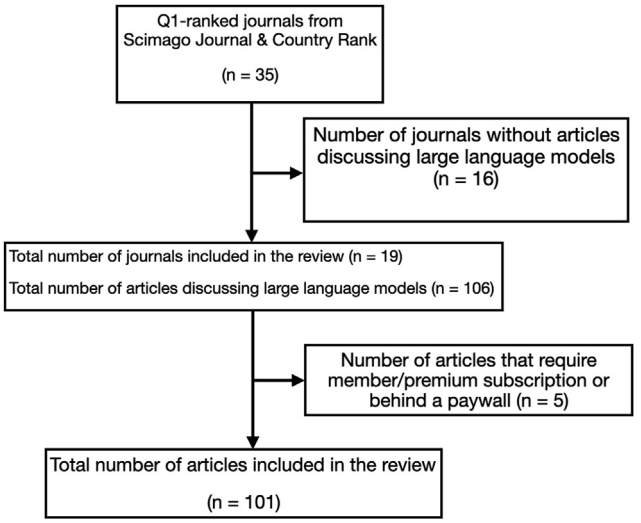


**Figure 1.** Diagram for identification of articles discussing large language models in the first quartile (Q1)-ranked journals from Scimago Journal & Country Rank.

Table 1. Overview of Articles Included in the Review

| Journal | Article Type | Title | Publication Date | First Author | Number of Listed Authors | Countries of Affiliation Listed Authors |
|---|---|---|---|---|---|---|
| Eye | Brief communication | Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot | 23-04 | Tsui et al[9] | 7 | USA |
| | Comment | Comparative analysis of large language models in the Royal College of Ophthalmologists fellowship exams | 23-12 | Raimondi et al[10] | 5 | UK, Italy |
| | Brief communication | Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination | 23-12 | Lin et al[11] | 5 | USA |
| | Correspondence | GPT-4 for triaging ophthalmic symptoms | 23-12 | Waisberg et al[12] | 7 | Ireland, USA |
| | Editorial | ChatGPT in ophthalmology: the dawn of a new era? | 23-06 | Ting et al[13] | 3 | UK, Singapore |
| | Brief communication | Modern threats in academia: evaluating plagiarism and artificial intelligence detection scores of ChatGPT | 24-02 | Taloni et al[14] | 3 | Italy |
| | Comment | GPT-4 to document ophthalmic postoperative complications | 24-02 | Waisberg et al[15] | 7 | Ireland, USA |
| | Brief communication | Large language models in vitreoretinal surgery | 24-03 | Anguita et al[16] | 4 | UK, Switzerland |
| | Comment | Large language model (LLM)-driven chatbots for neuro-ophthalmic medical education | 24-03 | Waisberg et al[17] | 4 | UK, USA |
| | Comment | Google's AI chatbot "Bard": a side-by-side comparison with ChatGPT and its utilization in ophthalmology | 24-03 | Waisberg et al[18] | 7 | UK, Ireland, USA |
| | Comment | How to use large language models in ophthalmology: from prompt engineering to protecting confidentiality | 24-03 | Kleinig et al[19] | 6 | Australia |
| | Comment | Comment on: "Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination" and "ChatGPT in ophthalmology: the dawn of a new era?" | 24-03 | Ghadiri[20] | 1 | UK |
| | Comment | ChatGPT to document ocular infectious diseases | 24-04 | Masalkhi et al[21] | 7 | Ireland, USA, UK |
| | Comment | Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments | 24-04 | Waisberg et al[22] | 7 | UK, USA |
| | Article | Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients | 24-05 | Cappellani et al[23] | 5 | USA |
| | Review article | ChatGPT and Beyond: An overview of the growing field of large language models and their use in ophthalmology | 24-05 | Kedia et al[24] | 4 | USA |
| | Comment | Google DeepMind's Gemini AI versus ChatGPT: a comparative analysis in ophthalmology | 24-06 | Masalkhi et al[25] | 4 | Ireland, UK, USA |
| | Comment | A side-by-side evaluation of Llama 2 by Meta with ChatGPT and its application in ophthalmology | 24-02 | Masalkhi et al[26] | 7 | Ireland, UK, USA |
| | Article | ChatGPT-3.5 and Bing Chat in ophthalmology: an updated evaluation of performance, readability, and informative sources | 24-03 | Tao et al[27] | 4 | Canada |
| | Article | Google Gemini and Bard artificial intelligence chatbot performance in ophthalmology knowledge assessment | 24-04 | Mihalache et al[28] | 8 | Canada |
| | Comment | Artificial intelligence chatbot interpretation of ophthalmic multimodal imaging cases | 24-04 | Mihalache et al[29] | 9 | Canada |
| | Comment | OpenAI's Sora in ophthalmology: revolutionary generative AI in eye health | 24-04 | Waisberg et al[30] | 4 | UK, USA |

(*Continued*)

Table 1. (Continued.)

| Journal | Article Type | Title | Publication Date | First Author | Number of Listed Authors | Countries of Affiliation Listed Authors |
|---|---|---|---|---|---|---|
| British Journal of Ophthalmology | Original research | Assessing the medical reasoning skills of GPT-4 in complex ophthalmology cases | 24-02 | Milad et al[31] | 13 | Canada, UK |
| | Original research | Capabilities of GPT-4 in ophthalmology: an analysis of model entropy and progress towards human-level medical question answering | 23-11 | Antaki et al[32] | 8 | UK, Canada |
| | Original research | Comparing generative and retrieval-based chatbots in answering patient questions regarding age-related macular degeneration and diabetic retinopathy | 24-05 | Cheong et al[33] | 16 | Singapore, China, UK, USA, South Korea |
| | Original research | Exploring AI chatbots' capability to suggest surgical planning in ophthalmology: ChatGPT versus Google Gemini analysis of retinal detachment cases | 24-03 | Carlà et al[34] | 8 | Italy |
| | Review | Foundation models in ophthalmology | 24-06 | Chia et al[35] | 6 | UK, Canada, Australia, USA |
| | Original research | ICGA-GPT: report generation and question answering for indocyanine green angiography images | 24-03 | Chen et al[36] | 7 | China |
| | Review | Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities | 24-05 | Sevgi et al[37] | 3 | UK |
| | Review | New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology | 22-07 | Nath et al[38] | 5 | Canada, UK, USA |
| | Original research | Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions | 23-11 | Fowler et al[39] | 3 | UK |
| | Systematic review | Review of emerging trends and projection of future developments in large language models research in ophthalmology | 23-12 | Wong et al[40] | 7 | UK, Singapore, Hong Kong |
| | Review | Towards regulatory generative AI in ophthalmology healthcare: a security and privacy perspective | 24-06 | Wang et al[41] | 5 | China |
| | Original research | Unveiling the clinical incapabilities: a benchmarking study of GPT-4V(ision) for ophthalmic multimodal image analysis | 24-05 | Xu et al[42] | 4 | China, Hong Kong |
| Ophthalmic and Physiological Optics | Original article | Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia | 23-07 | Biswas et al[43] | 5 | UK |
| | Review article | Utility of artificial intelligence-based large language models in ophthalmic care | 24-02 | Biswas et al[44] | 5 | UK |

Table 1. (Continued.)

| Journal | Article Type | Title | Publication Date | First Author | Number of Listed Authors | Countries of Affiliation Listed Authors |
|---|---|---|---|---|---|---|
| JAMA Ophthalmology | Brief report | Vision-Language Models for Feature Detection of Macular Diseases on Optical Coherence Tomography | 24-06 | Antaki et al[45] | 3 | UK |
| | Invited commentary | Large Language Models and the Shoreline of Ophthalmology | 24-04 | Young and Zhao[46] | 2 | USA |
| | Letters | Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research | 23-12 | Taloni et al[47] | 3 | Italy |
| | Brief report | Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment | 23-06 | Mihalache et al[48] | 3 | Canada |
| | Editorial | What Artificial Intelligence Chatbots Mean for Editors, Authors, and Readers of Peer-Reviewed Ophthalmic Literature | 23-06 | Bressler[49] | 1 | USA |
| | Letters | Performance of an Upgraded Artificial Intelligence Chatbot for Ophthalmic Knowledge Assessment | 23-08 | Mihalache et al[50] | 4 | Canada |
| | Invited commentary | Exploring the Test-Taking Capabilities of Chatbots - From Surgeon to Sommelier | 23-08 | Chia and Keane[51] | 2 | UK |
| | Original investigation | Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots | 23-09 | Hua et al[52] | 6 | USA |
| | Invited commentary | Chatbots, Artificial Intelligence, and the Future of Scientific Reporting | 23-09 | Volpe and Mirza[53] | 2 | USA |
| | Research letter | Accuracy of Vitreoretinal Disease Information From an Artificial Intelligence Chatbot | 23-09 | Caranfa et al[54] | 4 | USA |
| | Comment & response | Advances in Artificial Intelligence Chatbot Technology in Ophthalmology | 23-11 | Lin et al and Mihalache et al[55] | 3 and 3 | USA AND Canada |
| | Brief report | Assessment of a Large Language Model's Responses to Questions and Cases About Glaucoma and Retina Management | 23-04 | Huang et al[56] | 5 | USA |
| | Original investigation | Accuracy of an Artificial Intelligence Chatbot's Interpretation of Clinical Ophthalmic Images | 24-04 | Mihalache et al[57] | 12 | Canada |
| Graefe's Archive for Clinical and Experimental Ophthalmology | Brief communication | Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis | 23-10 | Rasmussen et al[58] | 4 | Denmark |
| | Brief communication | ChatGPT and scientific abstract writing: pitfalls and caution | 23-11 | Ali and Singh[59] | 2 | India |
| | Full paper | Diagnostic capabilities of ChatGPT in ophthalmology | 24-01 | Shemer et al[60] | 11 | Israel |
| | Full paper | Large language models as assistance for glaucoma surgical cases: a ChatGPT vs. Google Gemini comparison | 24-04 | Carlà et al[61] | 7 | Italy |

(*Continued*)

Table 1. (Continued.)

| Journal | Article Type | Title | Publication Date | First Author | Number of Listed Authors | Countries of Affiliation Listed Authors |
|---|---|---|---|---|---|---|
| American Journal of Ophthalmology | Perspective | Large Language Models in Ophthalmology Scientific Writing: Ethical Considerations Blurred Lines or Not at All? | 23-10 | Salimi and Saheb[62] | 2 | Canada |
| | Full-length article | Performance of Generative Large Language Models on Ophthalmology Board-Style Questions | 23-10 | Cai et al[63] | 7 | USA |
| | Correspondence | Comment on: Performance of Generative Large Language Models on Ophthalmology Board Style Questions | 23-12 | Kleebayoon and Wiwanitkit[64] | 2 | Cambodia, Nigeria |
| | Correspondence | Comment on: Large Language Models in Ophthalmology Scientific Writing: Ethical Considerations Blurred Lines or Not at All? | 24-02 | Metze et al[65] | 4 | Brazil |
| | Full-length article | Predicting glaucoma before onset using a large language model chatbot | 24-30 | Huang et al[66] | 7 | USA |
| | Correspondence | Reply to Comment on: Performance of Generative Large Language Models on Ophthalmology Board Style Questions | 24-12 | Cai and Alabiad[67] | 2 | USA |
| | Correspondence | Comment on: Large Language Models in Ophthalmology Scientific Writing: Ethical Considerations Blurred Lines or Not at All? | 24-02 | Jessup and Coroneo[68] | 2 | Australia |
| | Full-length article | Using Large Language Models to Generate Educational Materials on Childhood Glaucoma | 24-04 | Dihan et al[69] | 8 | USA |
| Ophthalmology Science | Research article | A Comparative Study of Responses to Retina Questions from Either Experts, Expert-Edited Large Language Models, or Expert-Edited Large Language Models Alone | 24-02 | Tailor et al[70] | 17 | USA |
| | Research article | A User-friendly Approach for the Diagnosis of Diabetic Retinopathy Using ChatGPT and Automated Machine Learning | 24-02 | Mohammadi and Nguyen[71] | 2 | USA |
| | Research article | Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of Its Successes and Shortcoming | 24-05 | Antaki et al[72] | 5 | Canada |
| | Review | Generative Artificial Intelligence Through ChatGPT and Other Large Language Models in Ophthalmology: Clinical Applications and Challenges | 23-09 | Tan et al[73] | 11 | Singapore, UK, USA |
| | Research article | Interpretation of Clinical Retinal Images Using an Artificial Intelligence Chatbot | 24-04 | Mihalache et al[74] | 12 | Canada |
| Current Opinion in Ophthalmology | Review | Applications of artificial intelligence-enabled robots and chatbots in ophthalmology: recent advances and future trends | 24-05 | Madadi et al[75] | 6 | USA |
| | Review | ChatGPT enters the room: what it means for patient counseling, physician education, academics, and disease management | 24-05 | Momenaei et al[76] | 7 | USA, Singapore |
| | Review | Vision of the future: large language models in ophthalmology | 24-05 | Tailor et al[77] | 4 | USA |
| Journal of Pediatric Ophthalmology & Strabismus | Correspondence | Prompt Engineering: Helping ChatGPT Respond Better to Patients and Parents | 24-02 | Chen and Granet[78] | 2 | USA |
| | Correspondence | Chatbot ChatGPT-4 and Frequently Asked Questions About Amblyopia and Childhood Myopia | 24-01 | Daungsuwapong and Wiwanitkit[79] | 2 | Lao People's Democratic Republic |
| | Reply to correspondence | Prompt Engineering: Helping ChatGPT Respond Better to Patients and Parents | 24-02 | Suh et al[80] | 4 | Iran |
| | Editorial | Pediatric Ophthalmology and Large Language Models: AI Has Arrived | 24-02 | Wagner[81] | 1 | USA |

Table 1. (Continued.)

| Journal | Article Type | Title | Publication Date | First Author | Number of Listed Authors | Countries of Affiliation Listed Authors |
|---|---|---|---|---|---|---|
| Ophthalmology Retina | Research article | Appropriateness and Readability of ChatGPT-4-Generated Responses for Surgical Treatment of Retinal Diseases | 23-10 | Momenaei et al[82] | 12 | USA |
| | Reports | Chatbot and Academy Preferred Practice Pattern Guidelines on Retinal Diseases | 24-03 | Mihalache et al[83] | 8 | Canada |
| | Correspondence | Re: Kianian et al.: Enhancing the assessment of large language models in medical information generation | 24-05 | Eleiwa and Elhusseiny[84] | 2 | Egypt, USA |
| | Correspondence | Re: Momenaei et al.: Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases | 23-10 | Bommakanti et al[85] | 4 | USA |
| | Reply to correspondence | Re: Kianian et al.: Enhancing the assessment of large language models in medical information generation | 24-05 | Kianian et al[86] | 4 | USA |
| | Reply to correspondence | Re: Momenaei et al.: Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases | 23-10 | Momenaei et al[87] | 12 | USA |
| | Research article | The Use of Large Language Models to Generate Education Materials about Uveitis | 24-02 | Kianian et al[88] | 4 | USA |
| Ophthalmology and Therapy | Brief report | What can GPT-4 do for Diagnosing Rare Eye Diseases? A Pilot Study | 23-09 | Hu et al[89] | 7 | China |
| | Original research | Artificial Intelligence-Based ChatGPT Responses for Patient Questions on Optic Disc Drusen | 23-09 | Potapenko et al[90] | 4 | Denmark |
| | Original research | The Use of ChatGPT to Assist in Diagnosing Glaucoma Based on Clinical Case Reports | 23-09 | Delsoz et al[91] | 7 | USA |
| | Letter | A Letter to the Editor Regarding "The Use of ChatGPT to Assist in Diagnosing Glaucoma Based on Clinical Case Reports" | 24-04 | Yaghy and Porteny[92] | 2 | USA |
| | Letter | A Response to: Letter to the Editor Regarding "The Use of ChatGPT to Assist in Diagnosing Glaucoma Based on Clinical Case Reports" | 24-04 | Delsoz et al[93] | 7 | USA |
| Journal of Glaucoma | Perspective article | Breaking Barriers in Behavioral Change: The Potential of AI-Driven Motivational Interviewing | 24-07 | Abid and Baxter[94] | 2 | USA |
| | Original study | Can ChatGPT Aid Clinicians in Educating Patients on the Surgical Management of Glaucoma | 24-02 | Kianian et al[95] | 3 | USA |
| | Case report/Brief report | Performance of ChatGPT on responding to common online questions regarding key information gaps in glaucoma | 24-02 | Wu et al[96] | 4 | USA |
| Survey of Ophthalmology | Review article | Generative artificial intelligence in ophthalmology | 24-01 | Waisberg et al[97] | 7 | UK, USA, Ireland |
| Clinical Ophthalmology | Original research | The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis | 24-02 | Raghu et al[98] | 6 | India |
| | Letter | The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis [Letter] | 24-02 | Fikri[99] | 1 | Indonesia |
| | Response to letter | The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis [Response to Letter] | 23-10 | Raghu et al[100] | 6 | India |

(*Continued*)

Table 1. (Continued.)

| Journal | Article Type | Title | Publication Date | First Author | Number of Listed Authors | Countries of Affiliation Listed Authors |
|---|---|---|---|---|---|---|
| Contact Lens and Anterior Eye | Research article | Are artificial intelligence chatbots a reliable source of information about contact lenses? | 24-03 | García-Porta et al[101] | 8 | Spain, UK, Ireland |
| | Research article | Assessing the proficiency of artificial intelligence programs in the diagnosis and treatment of cornea, conjunctiva, and eyelid diseases and exploring the advantages of each other's benefits | 24-05 | Sensoy and Citirik[102] | 2 | Turkey |
| Journal of Cataract & Refractive Surgery | From the editor | Artificial intelligence and academic publishing | 23-10 | Dupps, Jr.[103] | 1 | USA |
| | Letters | Comment on: Artificial intelligence chatbot and Academy Preferred Practice Pattern Guidelines on cataract and glaucoma | 24-05 | Daungsuwapong and Wiwanitkit[104] | 2 | India |
| | Letters | Reply: Artificial intelligence chatbot and Academy Preferred Practice Pattern Guidelines on cataract and glaucoma | 23-10 | Mihalache et al[105] | 4 | Canada |
| Retina | Original study | Performance assessment of an artificial intelligence chatbot in clinical vitreoretinal scenarios | 24-02 | Maywood et al[106] | 4 | USA |
| | Original study | The ability of artificial intelligence chatbots ChatGPT and Google Bard to accurately convey preoperative information for patients undergoing ophthalmic surgeries | 23-09 | Patil et al[107] | 11 | Canada |
| The Ocular Surface | Editorial | Readership awareness series - Paper 4: Chatbots and ChatGPT - Ethical considerations in scientific publications | 23-09 | Ali and Djalilian[108] | 2 | India |
| Ophthalmology | Editorial | The Pros and Cons of Artificial Intelligence Authorship in Ophthalmology | 23-09 | Van Gelder[109] | 1 | USA |

AI = artificial intelligence; NLP = natural language processing; UK = United Kingdom; USA = United States of America.

ophthalmology have only been around for about 2 years, its impact on research productivity has been arguably increasing, producing over 100 articles from top ophthalmology journals. An increase in team science collaboration has also been a key finding from this study; there was an average of 5.3 authors per article, and author affiliations contributing to LLM-related articles in ophthalmology were noted from different countries in the world. There was also a noted increase in the publication trends showcasing increased interest in the topic of LLM. One possible reason behind this is the availability of enhanced LLM versions or a new model to study or compare with different or older versions. Another possible reason is the exploration of the different models' utility in the field of ophthalmology, as discussed further in this section. Furthermore, there were noted discussions from published journals as noted with multiple published correspondence, comments, responses, and letters. This signifies the exchange of ideas from different authors and special interests in topics such as LLM.

Further parsing and analysis of the articles showed various models, with different versions of ChatGPT being the most used LLM. There were several implications regarding the use of LLM in ophthalmology. Generally, the different LLMs may benefit medical students, ophthalmology residents, ophthalmologists, researchers, and even patients or individuals seeking answers to basic or even advanced ophthalmologic queries. However, findings from different journals have also noted several drawbacks to its use.

## Medical Education

The performance and accuracy of LLM chatbots in answering ophthalmology examination questions show considerable variation. In general, these chatbots have demonstrated moderate to high accuracy in multiple test scenarios, including board examinations and clinical question assessments.[10,27,28,31,32,39,48,102] For example, ChatGPT models achieved differing accuracies depending on the complexity of the questions, with higher accuracies noted in simpler, text-based questions compared to more complex imaging-based or higher-order questions.[11] Bing and Bard chatbots also showed strong performance in text-based multiple-choice questions.[28] The legacy models showed lower accuracies, but newer versions like GPT Plus indicated better consistency and improved performance.[72] Moreover, performance improvements were observed in updated versions of the chatbots, which delivered more accurate responses across all question categories.[50] However, challenges remain, particularly in tasks involving image interpretation and multistep reasoning, where some models still offer significant rates of inaccuracies and nonlogical reasoning.[63,67,96]

The insights and comparisons of LLMs in addressing ophthalmology questions reveal several key findings. First, regarding the length of questions, different results have been published. Contrary to the abovementioned results, some publications show that the length of the prompt may not impact the performance of LLMs, indicating that the complexity or depth of questions might not be directly associated with their ability to provide accurate answers.[48] Also, ChatGPT tends to perform better when a higher percentage of human peers also get the right answer, suggesting it aligns with collective human understanding.[72] Therefore, several tools have been developed to assess the quality of health information provided by LLMs, such as DISCERN (an instrument to judge the quality of written consumer health information on treatment choices),[110] Patient Education Materials Assessment Tool, and the Education Quality Improvement Program. These tools are designed to evaluate the accuracy, comprehensiveness, and patient-centric nature of information, which can help improve the reliability of content generated by LLMs.[84] Furthermore, there are recommendations for LLM developers to adjust the reading level of responses, especially those containing medical information, to meet standards suggested by the American Medical Association, such as maintaining a sixth-grade reading level to ensure comprehensibility.[64]

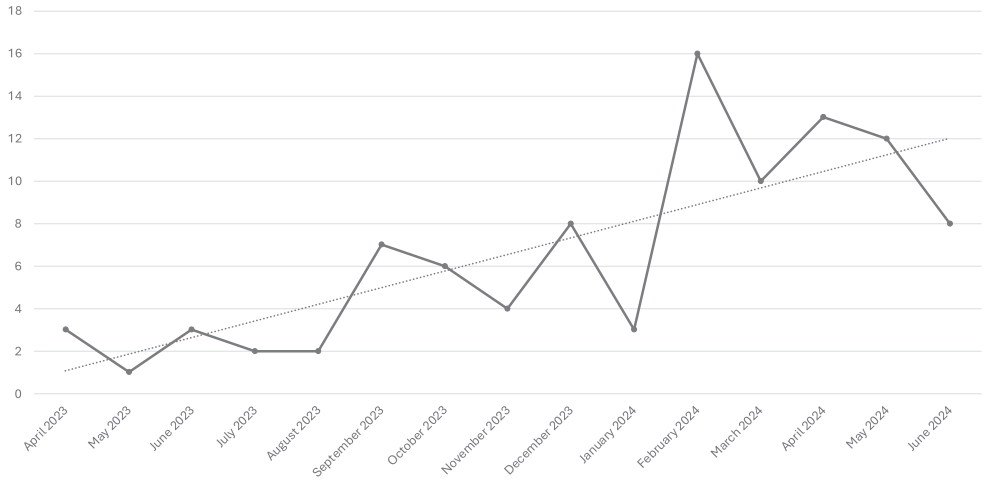


**Figure 2.** Monthly ophthalmology-related publications of large language models (LLMs) in the first quartile (Q1) ophthalmology journals. Note the increasing publication trend, which indicates an increase in research productivity since the availability of the use of LLM. Note that the earliest article discussing LLM in ophthalmology, published in July 2022, was not included in this figure.

Table 5. Article Background of Study About Large Language Model

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot | Tsui et al[9] | Evaluating chatbot answers to common ocular symptoms. | ChatGPT | Patient education |
| Comparative analysis of large language models in the Royal College of Ophthalmologists fellowship exams | Raimondi et al[10] | Evaluating performance of chatbot on Fellowship of Royal College of Ophthalmologists (FRCOphth) exam. | ChatGPT-3.5, ChatGPT-4, Bard, Bing | Medical education |
| Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination | Lin et al[11] | Evaluating performance of ChatGPT-3.5, ChatGPT-4, and human graders on ophthalmology exam. | ChatGPT-3.5, ChatGPT-4 | Medical education |
| GPT-4 for triaging ophthalmic symptoms | Waisberg et al[12] | Commenting on evaluation of chatbot answers to common ocular symptoms: prompts were imprecise. | ChatGPT-4 | Patient education |
| ChatGPT in ophthalmology: the dawn of a new era? | Ting et al[13] | Explaining conception of ChatGPT, evaluating roles in ophthalmology. | ChatGPT | Patient education, medical education, medical assistance, research |
| Modern threats in academia: evaluating plagiarism and artificial intelligence detection scores of ChatGPT | Taloni et al[14] | Evaluating and evading plagiarism and detection scores of ChatGPT-4.0 when rephrasing original scientific literature. | ChatGPT | Research |
| GPT-4 to document ophthalmic postoperative complications | Waisberg et al[15] | Evaluating ChatGPT-4's ability to detect intraoperative and postoperative complications. | ChatGPT-4 | Clinical assistance |
| Large language models in vitreoretinal surgery | Anguita et al[16] | Evaluating accuracy of outputs of 3 LLMs regarding vitreoretinal surgery. | ChatGPT-3.5, Bing, Docs-GPT Beta | Patient education |
| Large language model (LLM)-driven chatbots for neuro-ophthalmic medical education | Waisberg et al[17] | Evaluating the benefits of ChatGPT-4 in neuro-ophthalmology teaching. | ChatGPT-4 | Clinical assistance |
| Google's AI chatbot "Bard": a side-by-side comparison with ChatGPT and its utilization in ophthalmology | Waisberg et al[18] | Evaluating the performance of Bard and ChatGPT-3.5 in various prompts in ophthalmology. | Bard, ChatGPT-3.5 | Patient education |
| How to use large language models in ophthalmology: from prompt engineering to protecting confidentiality | Kleinig et al[19] | Describing the uses and limitations of LLM in ophthalmology. | ChatGPT-3, ChatGPT-4, PaLM-2, Llama-2 | Clinical assistance |
| Comment on: "Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination" and "ChatGPT in ophthalmology: the dawn of a new era?" | Ghadiri[20] | Commenting on prior studies regarding LLMs, particularly ChatGPT, and made suggestions on future use of LLMs | ChatGPT-3.5, ChatGPT-4 | Clinical assistance |
| ChatGPT to document ocular infectious diseases | Masalkhi et al[21] | Evaluating the potential of ChatGPT in managing and detecting infectious ophthalmologic diseases. | ChatGPT-4 | Patient education |
| Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments | Waisberg et al[22] | Introducing proprietary smart glasses (Ray-Ban) and the integration of images and conversion of visual information to speech through Meta AI. | N/A | Clinical assistance, patient education |
| Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients | Cappellani et al[23] | Evaluating the veracity of information provided by ChatGPT in ophthalmology. | ChatGPT-3.5 | Patient education |

Table 5. (Continued.)

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| ChatGPT and Beyond: An overview of the growing field of large language models and their use in ophthalmology | Kedia et al[24] | Exploring the principles of LLMs by presenting current studies concerning the application of LLMs in medicine. | ChatGPT-3, ChatGPT-3.5, ChatGPT-4, BioMedLM, DRAGON, BioLinkBERT, PubMedBERT, BioMegatron, BioGPT, T5, ClinicalBert, GatorTron, SciBERT, PaLM, MedPaLM | Medical education, research, clinical assistance |
| Google DeepMind's Gemini AI versus ChatGPT: a comparative analysis in ophthalmology | Masalkhi et al[25] | Comparing Gemini and ChatGPT's capabilities in answering patient-centered ophthalmic questions. | Google Gemini, ChatGPT-4 | Patient education |
| A side-by-side evaluation of Llama 2 by Meta with ChatGPT and its application in ophthalmology | Masalkhi et al[26] | Comparing Llama 2 and ChatGPT in answering patient-centered ophthalmic questions and fundus images. | Llama 2, ChatGPT-3.5 | Patient education |
| ChatGPT-3.5 and Bing Chat in ophthalmology: an updated evaluation of performance, readability, and informative sources | Tao et al[27] | Comparing the performance of publicly available LLMs (GPT-4, Bing Chat, ChatGPT-3.5) on American Academy of Ophthalmology (AAO) exam questions. | ChatGPT-3.5, Bing | Medical education |
| Google Gemini and Bard artificial intelligence chatbot performance in ophthalmology knowledge assessment | Mihalache et al[28] | Evaluating LLMs (Google Gemini, Bard) knowledge in ophthalmology. | Google Gemini, Bard | Medical education |
| Artificial intelligence chatbot interpretation of ophthalmic multimodal imaging cases | Mihalache et al[29] | Evaluating LLMs performance on open-ended queries in ophthalmology. | ChatGPT-4 | Clinical assistance |
| OpenAI's Sora in ophthalmology: revolutionary generative AI in eye health | Waisberg et al[30] | Describing the potential of Sora focusing on different areas in ophthalmology as clinician and patient. | Sora | Clinical assistance, patient education, medical education |
| Assessing the medical reasoning skills of GPT-4 in complex ophthalmology cases | Milad et al[31] | Evaluating ability of ChatGPT-4 of answering ophthalmologic questions in multilevel clinical scenarios. | ChatGPT-4 | Medical education |
| Capabilities of GPT-4 in ophthalmology: an analysis of model entropy and progress towards human-level medical question answering | Antaki et al[32] | Evaluating proficiency of ChatGPT-4 on BSCS and OphthoQuestions. | ChatGPT-3.5, ChatGPT-4 | Medical education |
| Comparing generative and retrieval-based chatbots in answering patient questions regarding age-related macular degeneration and diabetic retinopathy | Cheong et al[33] | Evaluating accuracy of generative and retrieval-based chatbots when asking patient questions regarding age-related macular degeneration (AMD) and diabetic retinopathy (DR). | ChatGPT-3.5, ChatGPT-4, Bard, OcularBERT | Patient education |
| Exploring AI-chatbots' capability to suggest surgical planning in ophthalmology: ChatGPT versus Google Gemini analysis of retinal detachment cases | Carlà et al[34] | Evaluating the analysis skills of ChatGPT-3.5, ChatGPT-4, and Google Gemini in retinal detachment (RD) cases and therapy recommendations. | ChatGPT-3.5, ChatGPT-4, Google Gemini | Clinical assistance |
| Foundation models in ophthalmology | Chia et al[35] | Providing insights on LLMs for use in ophthalmology research and daily practice. | LLMs, VLMs | Clinical assistance |
| ICGA-GPT: report generation and question answering for indocyanine green angiography images | Chen et al[36] | Developing bilingual ICGA report generator and responder to questions. | Llama 2-7b | Clinical assistance |

(*Continued*)

Table 5. (Continued.)

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities | Sevgi et al[37] | Explaining tuning of LLMs and work processes of custom GPT's. | Google Gemini, ChatGPT-4, Claude 2, LLaMA | Medical education |
| New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology | Nath et al[38] | Reviewing LLMs (ChatGPT-3) and discussing possible usage. | GPT 3 | Clinical assistance |
| Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions | Fowler et al[39] | Evaluating LLMs (Chat-GPT, Bard) performance on Part 1 Fellowship of Ophthalmologists (FRCOphth) exam questions. | ChatGPT-4, Bard | Medical education |
| Review of emerging trends and projection of future developments in large language models research in ophthalmology | Wong et al[40] | Reviewing current topics in LLM research and suggesting future fields of interest. | ChatGPT-3.5, ChatGPT-4, Bard, Bing | Clinical assistance, medical education, research |
| Towards regulatory generative AI in ophthalmology healthcare: a security and privacy perspective | Wang et al[41] | Evaluating security and privacy risks of LLMs regarding model-level and data-level. | LLM in general | Research |
| Unveiling the clinical incapabilities: a benchmarking study of GPT-4V(ision) for ophthalmic multimodal image analysis | Xu et al[42] | Evaluating ability of interpreting multimodal images using ChatGPT-4 V(ision). | GPT-4V(ision) | Clinical assistance |
| Assessing the utility of ChatGPT as an artificial intelligence-based large language model for information to answer questions on myopia | Biswas et al[43] | Evaluating information provided by LLMs on myopia regarding accuracy and quality. | ChatGPT-3.5 | Patient education |
| Utility of artificial intelligence-based large language models in ophthalmic care | Biswas et al[44] | Evaluating LLMs in patient care. | ChatGPT-3, ChatGPT-3.5, ChatGPT-4, Isabel Pro, Cornea specialist, Bing chat, WebMD symptom checker, ChatGPT-4V(ision) | Clinical assistance |
| Vision-Language Models for Feature Detection of Macular Diseases on Optical Coherence Tomography | Antaki et al[45] | Analyzing VLMs ability in diagnosing disease and generate treatment plans. | Gemini Pro vision-language model (VLM) | Clinical assistance |
| Large Language Models and the Shoreline of Ophthalmology | Young and Zhao[46] | Analyzing answers in real-world clinical cases. | ChatGPT-4, ChatGPT-3.5 | Clinical assistance |
| Large Language Model Advanced Data Analysis Abuse to Create a Fake Data Set in Medical Research | Taloni et al[47] | Analyzing ability of ChatGPT-4 to design artificial data for scientific research. | ChatGPT-4 | Research |
| Performance of an Artificial Intelligence Chatbot in Ophthalmic Knowledge Assessment | Mihalache et al[48] | Evaluating ChatGPT performance in ophthalmology exam questions. | ChatGPT-3 | Medical education |
| What Artificial Intelligence Chatbots Mean for Editors, Authors, and Readers of Peer-Reviewed Ophthalmic Literature | Bressler[49] | Reviewing contribution of AI in scientific literature. | ChatGPT | Research |
| Performance of an Upgraded Artificial Intelligence Chatbot for Ophthalmic Knowledge Assessment | Mihalache et al[50] | Evaluating performance of updated ChatGPT in examination questions. | ChatGPT-4 | Medical education |

Table 5. (Continued.)

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| Exploring the Test-Taking Capabilities of Chatbots - From Surgeon to Sommelier | Chia and Keane[51] | Commenting on prior studies regarding the use and limitations of ChatGPT. | ChatGPT-4 | Medical education |
| Evaluation and Comparison of Ophthalmic Scientific Abstracts and References by Current Artificial Intelligence Chatbots | Hua et al[52] | Evaluating ability of writing scientific abstracts in multiple ChatGPT versions. | ChatGPT-4, ChatGPT-3.5 | Research |
| Chatbots, Artificial Intelligence, and the Future of Scientific Reporting | Volpe and Mirza[53] | Evaluating limitations of LLMs. | LLM | Research |
| Accuracy of Vitreoretinal Disease Information From an Artificial Intelligence Chatbot | Caranfa et al[54] | Evaluating accuracy and repeatability on LLMs on patient questions. | ChatGPT-3 | Patient education |
| Advances in Artificial Intelligence Chatbot Technology in Ophthalmology | Lin et al and Mihalache et al[55] | Reviewing fundamental ethical questions in AI. | ChatGPT-4, ChatGPT-3.5 | Medical education |
| Assessment of a Large Language Model's Responses to Questions and Cases About Glaucoma and Retina Management | Huang et al[56] | Comparing LLM and fellowship-trained specialists' performance in evaluating clinical cases. | ChatGPT-4 | Clinical assistance, Patient education |
| Accuracy of an Artificial Intelligence Chatbot's Interpretation of Clinical Ophthalmic Images | Mihalache et al[57] | Evaluating LLMs accuracy of interpreting images. | ChatGPT-4 | Clinical assistance |
| Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis | Rasmussen et al[58] | Evaluating LLM answers to patients' requests. | ChatGPT-3 | Patient education |
| ChatGPT and scientific abstract writing: pitfalls and caution | Ali and Singh[59] | Evaluating ChatGPT's ability in academic writing. | ChatGPT | Research |
| Diagnostic capabilities of ChatGPT in ophthalmology | Shemer et al[60] | Evaluating ability of ChatGPT in diagnosing clinical cases. | ChatGPT-3.5 | Clinical assistance |
| Large language models as assistance for glaucoma surgical cases: a ChatGPT vs. Google Gemini comparison | Carlà et al[61] | Reviewing LLMs ability in diagnosing diseases and suggesting treatments. | ChatGPT-4, Google Gemini | Clinical assistance |
| Large Language Models in Ophthalmology Scientific Writing: Ethical Considerations Blurred Lines or Not at All? | Salimi and Saheb[62] | Discussing use and ethical concerns of LLMs in ophthalmology research. | ChatGPT-3.5 | Research |
| Performance of Generative Large Language Models on Ophthalmology Board-Style Questions | Cai et al[63] | Evaluating performance of LLMs on board-style ophthalmology questions. | ChatGPT-3.5, ChatGPT-4.0, Bing Chat | Medical education |
| Comment on: Performance of Generative Large Language Models on Ophthalmology Board Style Questions | Kleebayoon and Wiwanitkit[64] | Discussing ethical questions regarding LLMs. | ChatGPT-3.5, ChatGPT-4.0, Bing Chat | Medical education |
| Comment on: Large Language Models in Ophthalmology Scientific Writing: Ethical Considerations Blurred Lines or Not at All? | Metze et al[65] | Evaluating hallucinations of LLMs in science. | ChatGPT-3.5 | Research |
| Predicting glaucoma before onset using a large language model chatbot | Huang et al[66] | Evaluating LLMs in predicting progress in ocular hypertension to glaucoma. | ChatGPT-4.0 | Clinical assistance |

(*Continued*)

Table 5. (Continued.)

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| Reply to Comment on: Performance of Generative Large Language Models on Ophthalmology Board Style Questions | Cai and Alabiad[67] | Reply to Comment: Highlighting limitations of LLMs in answering ophthalmology questions. | ChatGPT-3.5, ChatGPT-4.0, Bing Chat | Medical education |
| Comment on: Large Language Models in Ophthalmology Scientific Writing: Ethical Considerations Blurred Lines or Not at All? | Jessup and Coroneo[68] | Reviewing shortcomings of datasets used to train LLMs. | ChatGPT-3.5 | Research |
| Using Large Language Models to Generate Educational Materials on Childhood Glaucoma | Dihan et al[69] | Evaluating patient education materials generated by LLMs. | ChatGPT-3.5, ChatGPT-4, Bard | Patient education |
| A Comparative Study of Responses to Retina Questions from Either Experts, Expert-Edited Large Language Models, or Expert-Edited Large Language Models Alone | Tailor et al[70] | Evaluating LLM responses to patient requests | ChatGPT-3.5, ChatGPT-4, Bard, Bing | Patient education |
| A User-friendly Approach for the Diagnosis of Diabetic Retinopathy Using ChatGPT and Automated Machine Learning | Mohammadi and Nguyen[71] | Evaluating ChatGPT and VertexAI ability to train LLMs and analyze output data. | ChatGPT-4, Vertex AI | Clinical assistance |
| Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of Its Successes and Shortcoming | Antaki et al[72] | Evaluating ChatGPT performance in medical education. | ChatGPT legacy, ChatGPT Plus | Medical education |
| Generative Artificial Intelligence Through ChatGPT and Other Large Language Models in Ophthalmology: Clinical Applications and Challenges | Tan et al[73] | Explaining abilities and challenges of LLMs in ophthalmology. | ChatGPT-1, ChatGPT-2, ChatGPT-3, ChatGPT-4 | Research, Clinical assistance |
| Interpretation of Clinical Retinal Images Using an Artificial Intelligence Chatbot | Mihalache et al[74] | Evaluating imaging processing abilities of LLMs. | ChatGPT-4 | Clinical assistance |
| Applications of artificial intelligence-enabled robots and chatbots in ophthalmology: recent advances and future trends | Madadi et al[75] | Evaluating abilities and future areas of application of LLMs. | ChatGPT-3.5, ChatGPT-4, BARD, LlaMA, Falcon, Cohere, PaLM, Claude | Clinical assistance |
| ChatGPT enters the room: what it means for patient counseling, physician education, academics, and disease management | Momenaei et al[76] | Reviewing possibilities, limitations, and ethical questions regarding the use of ChatGPT. | ChatGPT-3.5, ChatGPT-4 | Patient education, Medical education |
| Vision of the future: large language models in ophthalmology | Tailor et al[77] | Reviewing abilities and future areas of applications of LLMs in ophthalmology. | ChatGPT-3.5, ChatGPT-4 | Medical education, Clinical assistance |
| Prompt Engineering: Helping ChatGPT Respond Better to Patients and Parents | Chen and Granet[78] | Prompt engineering to improve LLM output. | ChatGPT-4 | Research |
| Chatbot ChatGPT-4 and Frequently Asked Questions About Amblyopia and Childhood Myopia | Daungsuwapong and Wiwanitkit[79] | Reviewing ability of LLM to answer questions about myopia and amblyopia. | ChatGPT-4 | Research |
| Prompt Engineering: Helping ChatGPT Respond Better to Patients and Parents | Suh et al[80] | Prompt engineering to improve LLM output. | ChatGPT-4 | Research |
| Pediatric Ophthalmology and Large Language Models: AI Has Arrived | Wagner[81] | Reviewing LLMs use in pediatric ophthalmology research. | ChatGPT-4 | Research |

Table 5. (Continued.)

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| Appropriateness and Readability of ChatGPT-4-Generated Responses for Surgical Treatment of Retinal Diseases | Momenaei et al[82] | Evaluating LLM output in ophthalmology regarding their comprehensibility. | ChatGPT-4 | Research |
| Chatbot and Academy Preferred Practice Pattern Guidelines on Retinal Diseases | Mihalache et al[83] | Evaluating ChatGPT's comments on Preferred Practice Pattern Guidelines. | ChatGPT-4, ChatGPT-3.5 | Research |
| Re: Kianian et al.: Enhancing the assessment of large language models in medical information generation | Eleiwa and Elhusseiny[84] | Criticizing publication on LLMs medical information generation due to lack of quality parameter. | ChatGPT-3.5, Bard | Medical education |
| Re: Momenaei et al.: Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases | Bommakanti et al[85] | Emphasizing future use of patients with health-related questions. | ChatGPT-4, ChatGPT-3.5 | Medical education |
| Re: Kianian et al.: Enhancing the assessment of large language models in medical information generation | Kianian et al[86] | Emphasizing need of quality assessment tools for LLM use in future. | ChatGPT-3.5, Bard | Medical education |
| Re: Momenaei et al.: Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases | Momenaei et al[87] | Evaluating improving performance in LLMs for future use in ophthalmology. | ChatGPT-4 | Research |
| The Use of Large Language Models to Generate Education Materials about Uveitis | Kianian et al[88] | Evaluating LLMs in creating educational information about uveitis. | ChatGPT-3.5, Bard | Medical education |
| What can GPT-4 do for Diagnosing Rare Eye Diseases? A Pilot Study | Hu et al[89] | Evaluating ChatGPT-4 performance in identifying rare eye diseases. | ChatGPT-4 | Medical education |
| Artificial Intelligence-Based ChatGPT Responses for Patient Questions on Optic Disc Drusen | Potapenko et al[90] | Evaluating LLMs responses on patient questions regarding optic disc drusen. | ChatGPT-4 | Patient education |
| The Use of ChatGPT to Assist in Diagnosing Glaucoma Based on Clinical Case Reports | Delsoz et al[91] | Evaluating ChatGPT's ability to diagnose glaucoma compared to human graders. | ChatGPT-3.5 | Clinical assistance |
| A Letter to the Editor Regarding "The Use of ChatGPT to Assist in Diagnosing Glaucoma Based on Clinical Case Reports" | Yaghy and Porteny[92] | Suggested that the model of ChatGPT was incapable of learning from user interactions at that point. | ChatGPT-3.5 | Clinical assistance |
| A Response to: Letter to the Editor Regarding "The Use of ChatGPT to Assist in Diagnosing Glaucoma Based on Clinical Case Reports" | Delsoz et al[93] | Replied that Reinforcement Learning from Human Feedback (RLHF) allows ChatGPT to learn from human feedback. Suggested approach of starting new medical cases in new sessions to overcome the possible bias due to RLHF. | ChatGPT-3.5 | Clinical assistance |
| Breaking Barriers in Behavioral Change: The Potential of AI-Driven Motivational Interviewing | Abid and Baxter[94] | Evaluating potential of ChatGPT for motivational interviewing. | ChatGPT | Clinical assistance |
| Can ChatGPT Aid Clinicians in Educating Patients on the Surgical Management of Glaucoma | Kianian et al[95] | Evaluating ChatGPT's performance on creating and critically analyzing health information. | ChatGPT | Patient education |

(*Continued*)

Table 5. (Continued.)

| Title | First Author | Purpose | LLM Used/Discussed/Mentioned | LLM Application |
|---|---|---|---|---|
| Performance of ChatGPT on responding to common online questions regarding key information gaps in glaucoma | Wu et al[96] | Analyzing ChatGPT's responses to clinical questions. | ChatGPT-4 | Medical education |
| Generative artificial intelligence in ophthalmology | Waisberg et al[97] | Reviewing the usage of GAN in ophthalmology. | GAN, ChatGPT-4 | Clinical assistance |
| The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis | Raghu et al[98] | Evaluating the ability of ChatGPT to predict diabetic retinopathy outcome. | ChatGPT-4 | Clinical assistance |
| The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis [Letter] | Fikri[99] | Suggested that LLM needs improvements with regard to privacy and security. | ChatGPT-4 | Clinical assistance |
| The Utility of ChatGPT in Diabetic Retinopathy Risk Assessment: A Comparative Study with Clinical Diagnosis [Response to Letter] | Raghu et al[100] | Suggested that ChatGPT should be trained with detailed knowledge of eye anatomy, physiology, and diseases to make it more useful. | ChatGPT-4 | Clinical assistance |
| Are artificial intelligence chatbots a reliable source of information about contact lenses? | García-Porta et al[101] | Evaluating accuracy of LLMs in replying to questions regarding contact lenses. | ChatGPT-3.5, Perplexity, OpenAI | Medical education |
| Assessing the proficiency of artificial intelligence programs in the diagnosis and treatment of cornea, conjunctiva, and eyelid diseases and exploring the advantages of each other's benefits | Sensoy and Citirik[102] | Evaluating LLMs clinical knowledge in corneal, conjunctival, and eyelid diseases. | ChatGPT-3.5, Bing, Bard | Medical education |
| Artificial intelligence and academic publishing | Dupps, Jr.[103] | Evaluating LLMs in its ability to write academic papers. | ChatGPT | Research |
| Comment on: Artificial intelligence chatbot and Academy Preferred Practice Pattern Guidelines on cataract and glaucoma | Daungsuwapong and Wiwanitkit[104] | Human users of LLM should follow a code since input equals output in LLM models. | ChatGPT | Research |
| Reply: Artificial intelligence chatbot and Academy Preferred Practice Pattern Guidelines on cataract and glaucoma | Mihalache et al[105] | Suggested that patients should be careful with the use of ChatGPT and a multidisciplinary team should formulate guidelines for proper use of chatbots. | ChatGPT | Research |
| Performance assessment of an artificial intelligence chatbot in clinical vitreoretinal scenarios | Maywood et al[106] | Evaluating LLM's accuracy in providing information about vitreoretinal cases. | ChatGPT-3.5 turbo | Clinical assistance |
| The ability of artificial intelligence chatbots ChatGPT and Google Bard to accurately convey preoperative information for patients undergoing ophthalmic surgeries | Patil et al[107] | Evaluating LLM's generation of preoperative information regarding accuracy and relevancy of information. | ChatGPT-4, Google Bard | Patient education |
| Readership awareness series - Paper 4: Chatbots and ChatGPT - Ethical considerations in scientific publications | Ali and Djalilian[108] | Emphasizing ethical challenges in the use of LLMs. | ChatGPT | Research |
| The Pros and Cons of Artificial Intelligence Authorship in Ophthalmology | Van Gelder[109] | Discussing the use of LLMs in scientific writing. | ChatGPT | Research |

AI = artificial intelligence; GAN = generative adversarial network; ICGA = indocyanine green angiography; LLM = large language model; NLP = natural language processing.

Geographical differences also affect the performance of LLMs, as demonstrated by the varying accuracies of chatbots in Spain and the United Kingdom. Significant discrepancies in accuracy were noted in these countries, with ChatGPT showing higher accuracy and less bias in certain settings. This highlights the importance of adapting LLMs to local legislations and user characteristics to enhance their effectiveness and accuracy across different regions.[101]

Large language models like ChatGPT offer a wide range of applications within ophthalmology. Customizable versions such as EyeTeacher enhance active learning by generating relevant questions, and EyeAssistant supports summarizing clinical guidelines and peer-reviewed research. Additionally, EyeAssistant is capable of summarizing the latest research and updates for geographic atrophy treatment.[37] Moreover, ChatGPT has demonstrated superior performance compared to historic examination takers in some publications, indicating its potential impact on medical education.[39] Large language models are capable of simplifying complex medical information based on how they are prompted, with ChatGPT overperforming Bard.[88] In particular, ChatGPT-4 shows promise as a consultation assisting tool. It can help patients and family physicians with referral suggestions and assist junior ophthalmologists in diagnosing rare eye diseases.[89] This emphasizes the potential of LLMs to support both educational and clinical aspects of ophthalmology.

Ophthalmologists need to stay updated on LLM systems as they continue to evolve and play a more significant role in medical care and decision-making.[11] Reviews highlight the importance of customized instructions and effective information retrieval when adapting GPT models for specific tasks in ophthalmology.[37] It is important to ensure AI systems are explainable and control challenges related to validation, computational demands, data acquisition, and accessibility.[10] As LLMs become more integrated into online learning and clinical practice, there is a growing need for ophthalmologists to become experienced in using these models. Future research should aim to develop open-access LLMs that are trained with accurate and reliable ophthalmology data. Despite advancements, GPT-4 does not currently surpass ophthalmology trainees in performance.[31] A cautious approach is essential when using LLMs, and results should not be generalized. Implementations of LLMs need to be robust, reliable, safe, and fair.[51] The models can exhibit bias, requiring proper management practices, and the responsibility for their use remains with humans.[64]

Enhanced prompting strategies can significantly improve the performance of ChatGPT-4, particularly in complex clinical scenarios.[31] Nevertheless, ethical challenges arise with the use of LLMs in ophthalmology, as shown by the rapid advancements from ChatGPT-3.5 to ChatGPT-4, which outperformed its predecessor within 5 months.[55] However, the creative abilities of LLMs can also result in "hallucinations," in which the model generates responses that sound plausible but are inaccurate.[32] This underscores the need to address critical aspects of individual privacy and LLM accountability.[37] More importantly, LLMs should not handle sensitive content without supervision.[64,102] Guidelines are crucial for ensuring that the development of LLM technologies in health care is both responsible and ethical.[67]

## Clinical Assistance

ChatGPT-3 and ChatGPT-4 have successfully shown great potential in addressing complex clinical issues. ChatGPT-4 created a management plan for specific medical complications,[15] though it missed including an evaluation for neuromyelitis optica.[17] It provided accurate insight for many open-ended questions in multimodal imaging cases, though not all responses were completely correct.[29] This can enhance the utilization of ophthalmic resources and clinic workflow, as well as assist in LLM solution development for clinicians.[38] Large language models demonstrated a moderate level of accuracy in diagnosing ocular diseases from various clinical scenarios and imaging,[44,57] performing better in non−image-based questions[57] and specific conditions like retinal and corneal cases but struggling with rare diseases and uveitis.[60] One group reported a substantial agreement of a generated bilingual system with ophthalmologists in generating indocyanine green angiography reports.[36] A ChatGPT-4 variant designed for vision tasks had mixed success in interpreting ocular images, with its best results in slit-lamp images but weaker performance in other images.[42] Overall, its responses were moderately consistent with human answers. ChatGPT's ability to forecast the progression from ocular hypertension to glaucoma a year before disease onset was fairly accurate, although it could not make predictions from a single visit.[66] ChatGPT's precision in diagnosing glaucoma using specific cases was comparable to or superior to senior ophthalmology residents. It provided a broader range of differential diagnoses and mostly listed potential causes of glaucoma.[91] In recent comparisons, the performance of LLMs was found to be superior in diagnosing and treating glaucoma compared to fellowship-trained specialists. In the field of retina, these models matched the performance of specialists and even exceeded them in terms of completeness. This suggests that LLM tools could serve as valuable diagnostic and therapeutic adjuncts.[56] Additionally, using AI-driven tools like ChatGPT for motivational interviewing shows the potential to significantly improve patient adherence.[94]

Another model, vision language model, demonstrated a limited capacity to detect pathological features, with its diagnostic capabilities and vision functions not yet meeting expert-level standards in ophthalmology.[45] Furthermore, ChatGPT struggled to differentiate between multiple clinical findings, with human residents and attendings still outperforming the LLMs.[60] The challenges identified include lack of accuracy and coherence, absence of real-time internet access, opaque functioning, replication of existing biases, data security concerns, and unclear legal status concerning medical records.[73] Despite some improvements in accuracy over time, the use of generative adversarial networks in clinical settings calls for additional research due to their potential to create fictitious features. Patients mostly consult LLMs for advice and information, while practitioners use them for support in clinical and

educational contexts.[73,97] Although it may be easier for specialists to discern hallucinations as they have extensive experience and clinical decision-making capabilities, it is a critical issue for patients relying on LLMs who will believe in its output regardless of the source.

When comparing ChatGPT and Gemini by analyzing case descriptions and recommending surgical options, ChatGPT aligns with specialists' opinions more frequently than Gemini, particularly performing better in simpler cases. Although Gemini shows less optimization for medical applications, both models demonstrate the potential for detailed analysis of medical records in glaucoma.[61] Overall, LLMs have the potential to enhance surgical precision. This includes improvements in robotic surgery and AI-powered video analysis for identifying subtle movements. A specific technology mentioned is the symmetrical threshold noise reduction combiner, which helps address issues like hand tremors during surgery.[75] Further, ChatGPT has shown significant proficiency in answering complex clinical scenarios related to vitreoretinal conditions, though it sometimes struggles to provide comprehensive responses.[106]

Large language models show promise in improving clinical assistance in eye care. They offer support in disease diagnostics, symptom assessment, and patient triage. Large language models also contribute to patient education by providing engaging information and assisting with the preparation of ophthalmic operative notes and responses to patient inquiries.[44] Patients themselves can use these models to check and supplement their doctor's evaluations by generating their own assessments and examining clinical notes. Large language models help manage the cognitive burden associated with checklists in busy outpatient settings and may aid in identifying medical errors.[46] They also demonstrate potential in interpreting various ophthalmic imaging modalities.[57] For conditions like glaucoma, an ideal AI tool would be a customized chatbot with expert-level performance that delivers humanized responses to diverse questions.[66] Moreover, advancements in deep learning and neural networks are expected to lead to more precise medical interventions and earlier disease detection.[57,71,75] However, the use of ChatGPT as a preliminary screening tool for diabetic retinopathy needs further development before it can be clinically used.[98]

## Research

Artificial intelligence chatbots are increasingly being used for research-related questions. The ethical concerns that arise from the use of AI in research have been discussed in a number of papers where it has been suggested that AI cannot meet the criteria for authorship since it cannot be held accountable for the content.[49] Many journals insist on declaring the use of AI for writing papers, and it has also been suggested that LLMs can be used in combination with human input with proper referencing.[62] A major concern for the use of AI in generating scientific content is its propensity to hallucinate. One study estimated a 30% hallucination rate with made-up references and a lack of nuanced decision-making.[53] Probable reasons are the inaccuracy, limitation, or lack thereof of training data.[65] However, many scientific publications are behind a paywall that restricts use for training these LLMs.[68] Ethical concerns also include being able to detect AI-generated content. Artificial intelligence output detectors often are unable to detect AI-generated abstracts[52] that receive a low plagiarism score even though LLM models plagiarize from unknown sources, especially if the training data set is not large enough.[59] Questions about intellectual property also arise, and a need for the scientific community to come together and establish guidelines has been suggested.[108,109] Suggested ways of improving the data generated by AI are prompt engineering to make data more accessible and "steerability" in ChatGPT-4. Steerability refers to the ability of a user to modify the LLM AI's tone and style by giving certain instructions to get a different answer.[78] Artificial intelligence in research thus has tremendous potential, but it raises ethical concerns around authorship, accountability, hallucinations due to data limitations, and detection of AI-generated content, necessitating community-established guidelines.

## Patient Education

Another area where LLM use has expanded and been studied greatly is patient education. When patients use LLMs for information about ophthalmology-related questions, it has been found that the information generated is accurate and detailed with ChatGPT-4 giving better responses than ChatGPT-3.[76] Patient education materials generated using LLMs have been found to provide accurate and high-quality information.[69] In one study, LLM alone was found to be better than experts at providing quality information along with an empathetic approach.[77] For patient education purposes, different LLMs have different strengths. For example, ChatGPT provides detailed answers, Bing provides verifiable references, Bard provides accurate responses, and Gemini can provide age-based recommendations when prompted.[16,18,25] They can accurately respond to questions regarding retinal surgeries with a very high success rate, although readability scores suggest that a college degree would be required to understand the content generated.[82] Information about glaucoma surgery was found to be often beyond the average patient's reading level, of subpar quality, and posing a risk of misunderstanding, while information provided by ChatGPT on infectious keratitis was found to be relevant and thorough.[21,95] Occasionally, inaccurate and harmful responses have been reported, and clinicians are well advised to guide the patients regarding this risk.[90] Interestingly, multilingual translations can help patients with information in countries where English is not the first language.[76] Although the accuracy of LLMs has increased over time, it is recommended that patients must not take medical advice from LLMs as factual, because the information provided may be out of date due to lack of real-time internet access, made-up references, and hallucinations.[73] All in all, ChatGPT and Bard have demonstrated efficacy in patient education within ophthalmology, offering precise and comprehensive information. These models sometimes outperform even experts in delivering quality content with an empathetic approach, although readability

remains a challenge, often requiring a college-level understanding. Despite their increasing accuracy, occasional inaccuracies and potentially harmful responses necessitate clinician oversight to mitigate risks.

### Introduction to Multimodal AI

The new version of ChatGPT, GPT-4o supports multimodal capabilities, allowing it to process and generate content that includes not only text but also images and other media.[111] Some of the findings in LLM-based AI models in our article might be useful as they act as a baseline for how these models might perform with additional multimodal capabilities. Incorporating images with text might enhance diagnostics by analyzing medical images in conjunction with textual information. Some of the ethical concerns studied in our review may need to be addressed while utilizing the image comprehension capabilities of this newer version, especially concerning Health Insurance Portability and Accountability Act of 1996 violations. Any uploaded image data (e.g., retinal images) that may be used for biometric identification of individuals is a potential risk from the viewpoint of breach of privacy and security.

There are several limitations in the present study. Articles were only reviewed from the Q1 journals from the SJR platform. This might exclude relevant studies published in high-quality journals not ranked as Q1, potentially leading to a biased overview of the literature. Only English-language journal articles were included in the study, as it is the only available literature in Q1-ranked journals, and this may exclude significant research published in other languages, limiting the comprehensiveness of the review. In addition, since the study focused only on ophthalmology journals, the review may be biased because related research is also published in comprehensive journals other than ophthalmology. Lastly, articles requiring a premium subscription or membership were excluded from the study. This exclusion might result in missing some key findings that were not reviewed.

In conclusion, in this comprehensive review, we noted the increase in research productivity in the field of ophthalmology regarding LLM use. We also highlighted the evolving role of LLMs and AI in ophthalmology, addressing their potential and challenges across various applications. Large language models demonstrate significant promise in medical education, clinical assistance, research, and patient education. However, the review also identifies critical limitations, including inconsistencies in performance, biases, hallucinations, and ethical concerns. Like many others, we emphasize the need for continuous improvements in AI models, multidisciplinary collaboration, and stringent ethical guidelines to ensure the safe and effective integration of LLMs in ophthalmology. Future research should focus on addressing these challenges to fully harness the capabilities of LLMs for advancing ophthalmic care and improving patient outcomes.

## Footnotes and Disclosures

Originally received: August 19, 2024.
Final revision: November 27, 2024.
Accepted: December 13, 2024.
Available online: December 17, 2024. Manuscript no. XOPS-D-24-00304.

[1] Jacobs Retina Center, University of California, San Diego, La Jolla, California.

[2] Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California, San Diego, La Jolla, California.

[3] Retina Care Hospital, Nagpur, India.

[4] Department of Ophthalmology, University Hospital Augsburg, Augsburg, Germany.

[5] Department of Ophthalmology and Visual Sciences, Philippine General Hospital, University of the Philippines Manila, Manila City, Philippines.

[6] International Eye Institute, St. Luke's Medical Center Global City, Taguig City, Philippines.

[7] Department of Ophthalmology, Massachusetts Eye and Ear, Boston, Massachusetts.

[8] Harvard Medical School, Department of Ophthalmology, Boston, Massachusetts.

[9] Division of Ophthalmology Informatics and Data Science, Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California, San Diego, La Jolla, California.

*A.P.A. and I.D.N. contributed equally as first author.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

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

Financial support was provided by the National Institutes of Health Bridge2AI Program (AI-READI) Grant OT2OD032644 (F.G.P.K).

Support for Open Access publication was provided by the University of California San Diego.

HUMAN SUBJECTS: No human subjects were included in this study. This study entailed reviewing published literature from ophthalmology-related journals. It did not involve human participation and, hence, did not require institutional review board approval.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Agnihotri, Nagel, Artiaga, Guevarra, Sosuan, Kalaw

Data collection: Agnihotri, Nagel, Artiaga, Guevarra, Sosuan, Kalaw

Analysis and interpretation: Agnihotri, Nagel, Artiaga, Guevarra, Sosuan, Kalaw

Obtained funding: N/A

Overall responsibility: Agnihotri, Nagel, Artiaga, Guevarra, Sosuan, Kalaw

Abbreviations and Acronyms:
**AI** = artificial intelligence; **LLMs** = large language models; **ML** = machine learning; **SJR** = Scimago Journal & Country Rank.

Keywords:
Large language models, Generative artificial intelligence, Chatbots, ChatGPT.

Correspondence:
Fritz Gerald P. Kalaw, MD, Division of Ophthalmology Informatics and Data Science, Shiley Eye Institute, University of California San Diego, 9415 Gilman Drive, La Jolla, CA 92093. E-mail: fkalaw@health.ucsd.edu.

## References


1. Kalaw FGP, Cavichini M, Zhang J, et al. Ultra-wide field and new wide field composite retinal image registration with AI-enabled pipeline and 3D distortion correction algorithm. *Eye*. 2023;38:1189–1195.
2. Ting DSW, Pasquale LR, Peng L, et al. Artificial intelligence and deep learning in ophthalmology. *Br J Ophthalmol*. 2019;103:167–175.
3. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*. 2016;316:2402.
4. Ting DSW, Cheung CYL, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211.
5. Stevenson CH, Hong SC, Ogbuehi KC. Development of an artificial intelligence system to classify pathology and clinical features on retinal fundus images. *Clin Exp Ophthalmol*. 2019;47:484–489.
6. Thirunavukarasu AJ, Ting DSJ, Elangovan K, et al. Large language models in medicine. *Nat Med*. 2023;29:1930–1940.
7. Fatani B. ChatGPT for future medical and dental research. *Cureus*. 2023;15:e37285.
8. Journal rankings on ophthalmology. https://www.scimagojr.com/journalrank.php?category=2731. Accessed June 7, 2024.
9. Tsui JC, Wong MB, Kim BJ, et al. Appropriateness of ophthalmic symptoms triage by a popular online artificial intelligence chatbot. *Eye*. 2023;37:3692–3693.
10. Raimondi R, Tzoumas N, Salisbury T, et al. Comparative analysis of large language models in the Royal College of Ophthalmologists fellowship exams. *Eye*. 2023;37:3530–3533.
11. Lin JC, Younessi DN, Kurapati SS, et al. Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination. *Eye*. 2023;37:3694–3695.
12. Waisberg E, Ong J, Zaman N, et al. GPT-4 for triaging ophthalmic symptoms. *Eye*. 2023;37:3874–3875.
13. Ting DSJ, Tan TF, Ting DSW. ChatGPT in ophthalmology: the dawn of a new era? *Eye*. 2024;38:4–7.
14. Taloni A, Scorcia V, Giannaccare G. Modern threats in academia: evaluating plagiarism and artificial intelligence detection scores of ChatGPT. *Eye*. 2024;38:397–400.
15. Waisberg E, Ong J, Masalkhi M, et al. GPT-4 to document ophthalmic post-operative complications. *Eye*. 2024;38:414–415.
16. Anguita R, Makuloluwa A, Hind J, Wickham L. Large language models in vitreoretinal surgery. *Eye*. 2024;38:809–810.
17. Waisberg E, Ong J, Masalkhi M, Lee AG. Large language model (LLM)-driven chatbots for neuro-ophthalmic medical education. *Eye*. 2024;38:639–641.
18. Waisberg E, Ong J, Masalkhi M, et al. Google's AI chatbot "Bard": a side-by-side comparison with ChatGPT and its utilization in ophthalmology. *Eye*. 2024;38:642–645.
19. Kleinig O, Gao C, Kovoor JG, et al. How to use large language models in ophthalmology: from prompt engineering to protecting confidentiality. *Eye*. 2024;38:649–653.
20. Ghadiri N. Comment on: 'Comparison of GPT-3.5, GPT-4, and human user performance on a practice ophthalmology written examination' and 'ChatGPT in ophthalmology: the dawn of a new era?'. *Eye*. 2024;38:654–655.
21. Masalkhi M, Ong J, Waisberg E, et al. ChatGPT to document ocular infectious diseases. *Eye*. 2024;38:826–828.
22. Waisberg E, Ong J, Masalkhi M, et al. Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. *Eye*. 2024;38:1036–1038.
23. Cappellani F, Card KR, Shields CL, et al. Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients. *Eye*. 2024;38:1368–1373.
24. Kedia N, Sanjeev S, Ong J, Chhablani J. ChatGPT and Beyond: an overview of the growing field of large language models and their use in ophthalmology. *Eye*. 2024;38:1252–1261.
25. Masalkhi M, Ong J, Waisberg E, Lee AG. Google DeepMind's gemini AI versus ChatGPT: a comparative analysis in ophthalmology. *Eye*. 2024;38:1412–1417.
26. Masalkhi M, Ong J, Waisberg E, et al. A side-by-side evaluation of Llama 2 by meta with ChatGPT and its application in ophthalmology. *Eye*. 2024;38:1789–1792.
27. Tao BKL, Hua N, Milkovich J, Micieli JA. ChatGPT-3.5 and Bing Chat in ophthalmology: an updated evaluation of performance, readability, and informative sources. *Eye*. 2024;38:1897–1902.
28. Mihalache A, Grad J, Patil NS, et al. Google Gemini and Bard artificial intelligence chatbot performance in ophthalmology knowledge assessment. *Eye*. 2024;38:2530–2535.
29. Mihalache A, Huang RS, Cruz-Pimentel M, et al. Artificial intelligence chatbot interpretation of ophthalmic multimodal imaging cases. *Eye*. 2024;38:2491–2493.
30. Waisberg E, Ong J, Masalkhi M, Lee AG. OpenAI's Sora in medicine: revolutionary advances in generative artificial intelligence for healthcare. *Ir J Med Sci*. 2024;193:2105–2107.
31. Milad D, Antaki F, Milad J, et al. Assessing the medical reasoning skills of GPT-4 in complex ophthalmology cases. *Br J Ophthalmol*. 2024;108:1398–1405.
32. 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*. 2023;108:1371–1378.
33. 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;108:1343–1349.
34. Carlà MM, Gambini G, Baldascino A, et al. Exploring AI-chatbots' capability to suggest surgical planning in ophthalmology: ChatGPT versus Google Gemini analysis of retinal detachment cases. *Br J Ophthalmol*. 2024;108:1357–1369.
35. Chia MA, Antaki F, Zhou Y, et al. Foundation models in ophthalmology. *Br J Ophthalmol*. 2024;108:1341–1348.
36. Chen X, Zhang W, Zhao Z, et al. ICGA-GPT: report generation and question answering for indocyanine green angiography images. *Br J Ophthalmol*. 2024;108:1450–1456.
37. Sevgi M, Antaki F, Keane PA. Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities. *Br J Ophthalmol*. 2024;108:1354–1361.
38. Nath S, Marie A, Ellershaw S, et al. New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. *Br J Ophthalmol*. 2022;106:889–892.
39. Fowler T, Pullen S, Birkett L. Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions. *Br J Ophthalmol*. 2023;108:1349–1353.

40. Wong M, Lim ZW, Pushpanathan K, et al. Review of emerging trends and projection of future developments in large language models research in ophthalmology. *Br J Ophthalmol*. 2023;108:1362−1370.
41. Wang Y, Liu C, Zhou K, et al. Towards regulatory generative AI in ophthalmology healthcare: a security and privacy perspective. *Br J Ophthalmol*. 2024;108:1349−1353.
42. Xu P, Chen X, Zhao Z, Shi D. Unveiling the clinical incapabilities: a benchmarking study of GPT-4V(ision) for ophthalmic multimodal image analysis. *Br J Ophthalmol*. 2024;108:1384−1389.
43. 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 Physiologic Optic*. 2023;43:1562−1570.
44. Biswas S, Davies LN, Sheppard AL, et al. Utility of artificial intelligence-based large language models in ophthalmic care. *Ophthalmic Physiol Opt*. 2024;44:641−671. https://doi.org/10.1111/opo.13284.
45. Antaki F, Chopra R, Keane PA. Vision-Language models for feature detection of macular diseases on optical coherence tomography. *JAMA Ophthalmol*. 2024;142:573−576.
46. Young BK, Zhao PY. Large Language models and the shoreline of ophthalmology. *JAMA Ophthalmol*. 2024;142:375−376.
47. Taloni A, Scorcia V, Giannaccare G. Large Language model advanced data analysis abuse to create a fake data set in medical research. *JAMA Ophthalmol*. 2023;141:1174−1175.
48. Mihalache A, Popovic MM, Muni RH. Performance of an artificial intelligence chatbot in ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023;141:589.
49. Bressler NM. What artificial intelligence chatbots mean for editors, authors, and readers of peer-reviewed ophthalmic literature. *JAMA Ophthalmol*. 2023;141:514.
50. Mihalache A, Huang RS, Popovic MM, Muni RH. Performance of an upgraded artificial intelligence chatbot for ophthalmic knowledge assessment. *JAMA Ophthalmol*. 2023;141:798.
51. Chia MA, Keane PA. Exploring the test-taking capabilities of chatbots—from surgeon to sommelier. *JAMA Ophthalmol*. 2023;141:800−801.
52. 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.
53. Volpe NJ, Mirza RG. Chatbots, artificial intelligence, and the future of scientific reporting. *JAMA Ophthalmol*. 2023;141:824.
54. Caranfa JT, Bommakanti NK, Young BK, Zhao PY. Accuracy of vitreoretinal disease information from an artificial intelligence chatbot. *JAMA Ophthalmol*. 2023;141:906.
55. Lin JC, Kurapati SS, Scott IU. Advances in artificial intelligence chatbot technology in ophthalmology. *JAMA Ophthalmol*. 2023;141:1088.
56. Huang AS, Hirabayashi K, Barna L, et al. Assessment of a Large Language Model's responses to questions and cases about glaucoma and retina management. *JAMA Ophthalmol*. 2024;142:371−375.
57. Mihalache A, Huang RS, Popovic MM, et al. Accuracy of an artificial intelligence chatbot's interpretation of clinical ophthalmic images. *JAMA Ophthalmol*. 2024;142:321.
58. 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;261:3041−3043.
59. Ali MJ, Singh S. ChatGPT and scientific abstract writing: pitfalls and caution. *Graefes Arch Clin Exp Ophthalmol*. 2023;261:3205−3206.
60. Shemer A, Cohen M, Altarescu A, et al. Diagnostic capabilities of ChatGPT in ophthalmology. *Graefes Arch Clin Exp Ophthalmol*. 2024;262:2345−2352.
61. Carlà MM, Gambini G, Baldascino A, et al. Large language models as assistance for glaucoma surgical cases: a ChatGPT vs. Google Gemini comparison. *Graefes Arch Clin Exp Ophthalmol*. 2024;262:2945−2959.
62. Salimi A, Saheb H. Large Language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am J Ophthalmol*. 2023;254:177−181.
63. Cai LZ, Shaheen A, Jin A, et al. Performance of generative Large Language models on ophthalmology board−style questions. *Am J Ophthalmol*. 2023;254:141−149.
64. Kleebayoon A, Wiwanitkit V. Comment on: performance of generative Large Language models on ophthalmology board style questions. *Am J Ophthalmol*. 2023;256:200.
65. Metze K, Lorand-Metze I, Morandin-Reis RC, Florindo JB. Comment on: large Language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am J Ophthalmol*. 2024;264:241−242.
66. Huang X, Raja H, Madadi Y, et al. Predicting glaucoma before onset using a large language model chatbot. *Am J Ophthalmol*. 2024;226:289−299.
67. Cai LZ, Alabiad C. Reply to comment on: performance of generative Large Language models on ophthalmology board style questions. *Am J Ophthalmol*. 2023;256:201.
68. Jessup AJC, Coroneo MT. Comment on: large Language models in ophthalmology scientific writing: ethical considerations blurred lines or not at all? *Am J Ophthalmol*. 2024;264:239−240.
69. Dihan Q, Chauhan MZ, Eleiwa TK, et al. Using Large Language models to generate educational materials on childhood glaucoma. *Am J Ophthalmol*. 2024;265:28−38.
70. Tailor PD, Dalvin LA, Chen JJ, et al. A comparative study of responses to retina questions from either experts, expert-edited Large Language models, or expert-edited Large Language models alone. *Ophthalmol Sci*. 2024;4:100485.
71. Mohammadi SS, Nguyen QD. A user-friendly approach for the diagnosis of diabetic retinopathy using ChatGPT and automated machine learning. *Ophthalmol Sci*. 2024;4:100495.
72. 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.
73. Tan TF, Thirunavukarasu AJ, Campbell JP, et al. Generative artificial intelligence through ChatGPT and other Large Language models in ophthalmology. *Ophthalmol Sci*. 2023;3:100394.
74. Mihalache A, Huang RS, Mikhail D, et al. Interpretation of clinical retinal images using an artificial intelligence chatbot. *Ophthalmol Sci*. 2024;4:100556.
75. Madadi Y, Delsoz M, Khouri AS, et al. Applications of artificial intelligence-enabled robots and chatbots in ophthalmology: recent advances and future trends. *Curr Opin Ophthalmol*. 2024;35:238−243.
76. Momenaei B, Mansour HA, Kuriyan AE, et al. ChatGPT enters the room: what it means for patient counseling, physician education, academics, and disease management. *Curr Opin Ophthalmol*. 2024;35:205−209.
77. Tailor PD, D'Souza HS, Li H, Starr MR. Vision of the future: large language models in ophthalmology. *Curr Opin Ophthalmol*. 2024;35:391−402.

78. Chen JS, Granet DB. Prompt engineering: helping ChatGPT respond better to patients and parents. *J Pediatr Ophthalmol Strabismus*. 2024;61:148−149.
79. Daungsupawong H, Wiwanitkit V. Chatbot ChatGPT-4 and frequently asked questions about amblyopia and childhood myopia. *J Pediatr Ophthalmol Strabismus*. 2024;61:151.
80. Suh DW, Nikdel M, Tavakoli M, Ghadimi H. Reply: prompt engineering: helping ChatGPT respond better to patients and parents. *J Pediatr Ophthalmol Strabismus*. 2024;61:149−150.
81. Wagner RS. Pediatric ophthalmology and Large Language models: AI has arrived. *J Pediatr Ophthalmol Strabismus*. 2024;61:80.
82. 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;7:862−868.
83. Mihalache A, Huang RS, Patil NS, et al. Chatbot and academy preferred practice pattern guidelines on retinal diseases. *Ophthalmol Retina*. 2024;8:723−725.
84. Eleiwa TK, Elhusseiny AM. Re: Kianian Enhancing the assessment of large language models in medical information generation (Ophthalmol Retina. 2024;8:195-201). *Ophthalmol Retina*. 2024;8:e15.
85. Bommakanti N, Caranfa JT, Young BK, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases (Ophthalmol Retina. 2023:7:862-868). *Ophthalmol Retina*. 2024;8:e1.
86. Kianian R, Sun D, Crowell EL, Tsui E. Reply. *Ophthalmol Retina*. 2024;8:e15−e16.
87. Momenaei B, Wakabayashi T, Shahlaee A, et al. Reply. *Ophthalmol Retina*. 2024;8:e1−e2.
88. Kianian R, Sun D, Crowell EL, Tsui E. The use of Large Language models to generate education materials about uveitis. *Ophthalmol Retina*. 2024;8:195−201.
89. Hu X, Ran AR, Nguyen TX, et al. What can GPT-4 do for diagnosing rare eye diseases? A pilot study. *Ophthalmol Ther*. 2023;12:3395−3402.
90. Potapenko I, Malmqvist L, Subhi Y, Hamann S. Artificial intelligence-based ChatGPT responses for patient questions on optic disc drusen. *Ophthalmol Ther*. 2023;12:3109−3119.
91. Delsoz M, Raja H, Madadi Y, et al. The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther*. 2023;12:3121−3132.
92. Yaghy A, Porteny JR. A letter to the editor regarding "the use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports.". *Ophthalmol Ther*. 2024;13:1813−1815.
93. Delsoz M, Raja H, Madadi Y, et al. A response to: letter to the editor regarding "the use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports.". *Ophthalmol Ther*. 2024;13:1817−1819.
94. Abid A, Baxter SL. Breaking barriers in behavioral change: the potential of AI-driven motivational interviewing. *J Glaucoma*. 2024;33:473−477.
95. Kianian R, Sun D, Giaconi J. Can ChatGPT aid clinicians in educating patients on the surgical management of glaucoma? *J Glaucoma*. 2024;33:94−100.
96. Wu JH, Nishida T, Moghimi S, Weinreb RN. Performance of ChatGPT on responding to common online questions regarding key information gaps in glaucoma. *J Glaucoma*. 2024;33:e54−e56.
97. Waisberg E, Ong J, Kamran SA, et al. Generative artificial intelligence in ophthalmology. *Surv Ophthalmol*. 2024;70:1−11.
98. Raghu K, T S, S Devishamani C, et al. The utility of ChatGPT in diabetic retinopathy risk assessment: a comparative study with clinical diagnosis. *Clin Ophthalmol*. 2023;17:4021−4031.
99. Fikri E. The utility of ChatGPT in diabetic retinopathy risk assessment: a comparative study with clinical diagnosis [letter]. *OPTH*. 2024;18:127−128.
100. Raghu K, S T, S Devishamani C, et al. The utility of ChatGPT in diabetic retinopathy risk assessment: a comparative study with clinical diagnosis [response to letter]. *OPTH*. 2024;18:313−314.
101. García-Porta N, Vaughan M, Rendo-González S, et al. Are artificial intelligence chatbots a reliable source of information about contact lenses? *Contact Lens Anterior Eye*. 2024;47:102130.
102. Sensoy E, Citirik M. Assessing the proficiency of artificial intelligence programs in the diagnosis and treatment of cornea, conjunctiva, and eyelid diseases and exploring the advantages of each other benefits. *Contact Lens Anterior Eye*. 2024;47:102125.
103. Dupps WJ. Artificial intelligence and academic publishing. *J Cataract Refract Surg*. 2023;49:655−656.
104. Daungsupawong H, Wiwanitkit V. Comment on: artificial intelligence chatbot and academy preferred practice Pattern® guidelines on cataract and glaucoma. *J Cataract Refract Surg*. 2024;50:661−662.
105. Mihalache A, Huang RS, Popovic MM, Muni RH. Reply: artificial intelligence chatbot and academy preferred practice Pattern guidelines on cataract and glaucoma. *J Cataract Refract Surg*. 2024;50:662−663.
106. Maywood MJ, Parikh R, Deobhakta A, Begaj T. Performance assessment of an artificial intelligence chatbot in clinical vitreoretinal scenarios. *Retina*. 2024;44:954−964.
107. Patil NS, Huang R, Mihalache A, et al. THE ability of artificial intelligence chatbots ChatGPT and google bard to accurately convey preoperative information for patients undergoing ophthalmic surgeries. *Retina*. 2024;44:950−953.
108. Ali MJ, Djalilian A. Readership awareness series − paper 4: chatbots and ChatGPT - ethical considerations in scientific publications. *Ocul Surf*. 2023;28:153−154.
109. Van Gelder RN. The pros and cons of artificial intelligence authorship in ophthalmology. *Ophthalmology*. 2023;130:670−671.
110. Shan Y, Xing Z, Dong Z, et al. Translating and adapting the DISCERN instrument into a simplified Chinese version and validating its reliability: development and usability study. *J Med Internet Res*. 2023;25:e40733.
111. Stoner R. Unveiling ChatGPT-4o: next-gen features and their transformative impact. Unite.AI. https://www.unite.ai/unveiling-chatgpt-4o-next-gen-features-and-their-transformative-impact/; 2024. Accessed August 2, 2024.