



Large Language Models in Ophthalmology: A Bibliographic Analysis

© Neslihan Dilruba Köseoğlu¹, © T.Y. Alvin Liu²

¹Tufts Medical Center, Clinic of Ophthalmology, Boston, MA, USA

²Wilmer Eye Institute, School of Medicine, Johns Hopkins University, Baltimore, MD, USA

Abstract

This study evaluated the distribution of research on the use of large language models (LLMs) in ophthalmology through a bibliographic analysis of articles retrieved from PubMed through November 2024. Studies were categorized into four main areas of LLM application: clinical decision-making (further divided according to subspecialties), education, patient interactions, and miscellaneous applications. Descriptive statistics were used to analyze the distribution of studies by ophthalmic subspecialty, geographical region, journal quality, and author characteristics, including gender and scholarly impact (h-index and i10-index). The findings revealed that clinical decision-making was the most common application (43.7%), with the majority of studies in this subgroup focusing on the retina (39.5%). Geographically, most of the research originated from North America (48.3%), followed by Asia (29.9%) and Europe (20.7%). Most studies were published in high-impact journals (Q1 journals: 74.7%), particularly for those related to clinical decision-making in retina (80.0%), glaucoma (100%), and multiple subspecialties (87.5%). Gender disparities were evident across all author roles, with female authors accounting for only 29.9% of first authors, 25.3% of last authors, and 26.4% of corresponding authors. The results suggest a need for greater diversity in terms of gender and geographic representation in LLM research in ophthalmology to promote inclusive progress in the field.

Keywords: Large language models, ophthalmology, bibliographical analysis

Introduction

In recent years, artificial intelligence (AI), particularly in the form of large language models (LLMs), has revolutionized many aspects of science and become an integral part of research.¹ Models such as ChatGPT, BERT, and LLaMA utilize millions of parameters and draw from diverse data sources, including books, articles, and other text-based materials, to generate human-like text responses.² These LLMs differ architecturally: ChatGPT uses autoregressive transformers focused on generative language tasks; BERT employs bidirectional transformers for contextual understanding; and LLaMA has a standard decoder-only transformer architecture with minor adaptations, avoiding mixture-of-experts models to maximize training stability.^{3,4,5} Despite these differences, all are commonly used for generating responses.

While AI has been utilized in medicine for decades, LLMs are now transforming diagnostics through image analysis and enhancing clinical decision-making by efficiently processing vast amounts of data, clinical information, and patient records, paving the way for precision medicine.⁶ Over the past few years, LLMs have become increasingly adopted across various medical specialties, including radiology, internal medicine, pediatrics, cardiovascular medicine, and many others.^{7,8,9} Research related to LLMs has increased significantly since 2023, reflecting a growing interest in their capabilities and applications. Ophthalmology, in particular, has been at the forefront of AI-related research, with growing interest in leveraging LLMs to advance patient care.¹⁰

However, LLM-related research also faces several limitations and challenges. First, LLMs are predominantly trained on English-language data, which may affect the reliability of their responses when applied to non-English

Cite this article as: Köseoğlu ND, Liu TYA. Large Language Models in Ophthalmology: A Bibliographic Analysis. *Turk J Ophthalmol.* 2026;56:119-130

Address for Correspondence: Neslihan Dilruba Köseoğlu, Tufts Medical Center, Clinic of Ophthalmology, Boston, MA, USA

E-mail: dilruba33@yahoo.com

ORCID-ID: orcid.org/0000-0001-9268-1461

Received: 16.04.2025

Revision Requested: 06.06.2025

Last Revision Received: 01.12.2025

Accepted: 01.01.2026

Publication Date: 27.04.2026

DOI: 10.4274/tjo.galenos.2026.79484



languages. As a result, research is often concentrated in regions where English is widely spoken. Regional disparities are evident across all AI studies, with the majority of publications and grants dominated by the US and China, which together have been shown to account for approximately 50% of all publications from 2014 to 2023.¹¹ Additionally, the AI Index Report (2024) showed that novel AI models developed in 2023 primarily originated in the US, followed by China and Europe, further highlighting the geographical gap in AI research.¹² This unequal distribution in research output is accompanied by other disparities within the field, including gender representation. The latest Global Gender Gap Report (2024) by the World Economic Forum revealed that only 22% of professionals in AI are women, a gender gap which is reflected in AI-related research.¹³

Given these known disparities in AI-related research, this study aimed to provide a closer look at LLM-related studies in ophthalmology. We performed a bibliographic analysis to evaluate the distribution of research on LLMs in ophthalmology, focusing on gender, academic impact, and the geographical origin of the studies.

Methods

A comprehensive literature search was conducted on PubMed to identify studies involving the use of LLMs in ophthalmology published up until November 2024. The search terms used were *large language models* AND *ophthalmology* OR *retina* OR *glaucoma* OR *cornea* OR *uvea* OR *pediatric ophthalmology* OR *neuro-ophthalmology*. The search results yielded a total of 194 studies. Only original investigations were included in the final analysis. Studies not related to LLMs or ophthalmology, review articles, meta-analyses, and commentaries were excluded from the final list.

Included studies were categorized into four groups based on their primary focus and application:

- **Clinical Decision-Making Applications:** Studies that investigated the use of LLMs in clinical decision-making and assessed LLM responses in diagnosing, managing, or providing clinical support for ophthalmic conditions. This category was further divided into ophthalmic subspecialties including retina, glaucoma, cornea/anterior segment, uveitis, neuro-ophthalmology, and pediatrics. Studies that encompassed more than one ophthalmic subspecialty or general ophthalmology were categorized as “multiple subspecialties.”

- **Educational Applications:** Studies that explored the use of LLMs in educational contexts, including their application in answering board-style questions and

developing educational materials for both patients and clinicians.

- **Patient Interaction Applications:** Studies that focused on how LLMs are used to improve patient communication and respond to frequently asked questions about ocular health, such as in the format of a chatbot.

- **Miscellaneous Applications:** Studies that explored other applications of LLMs in ophthalmology that do not neatly fit into other categories, particularly those that are highly technical in nature.

The initial categorization of studies was conducted by the first author and reviewed by the senior author. Discrepancies were resolved by discussion and consensus.

Descriptive statistics were used to analyze the distribution of studies by application type, ophthalmic subspecialty, geographical region, journal quality, and author characteristics. The geographic regions were determined based on the United Nations World Population Prospects, which divides the world into six continental regions: Africa, Asia, Europe, Latin America and the Caribbean, North America, and Oceania.¹⁴ Geographical region was assigned according to the first author’s affiliated institution. Journal quality was classified into quartiles Q1, Q2, Q3, and Q4 as retrieved from SCImago. Author characteristics included gender and scholarly impact. Gender information was obtained from institutional websites or professional profiles. Scholarly impact was assessed using the h-index and i10-index, both retrieved from Google Scholar. The h-index is a metric that measures the scientific impact of an author’s publications, defined as the maximum value of h for which the author has published at least h papers, each of which has been cited no fewer than h times.¹⁵ In contrast, the i10-index is a related metric used by Google Scholar that counts the number of publications with at least 10 citations.¹⁶

Results

Overall Results

A total of 87 original research studies were included in the final evaluation. In terms of application type, the most common was clinical decision-making ($n=38$, 43.7%), followed by educational applications ($n=22$, 25.3%), patient interaction applications ($n=18$, 20.7%), and miscellaneous applications ($n=9$, 10.3%) (Table 1). Within the clinical decision-making category, studies related to the retina were most prevalent, accounting for 15 studies (39.5%), while pediatrics had the least, with only 1 study (2.6%). Notably, research encompassing multiple subspecialties accounted for 21.1% of the studies in this group ($n=8$). There were

moderate contributions from glaucoma, cornea/anterior segment, uveitis, and neuro-ophthalmology. The number of studies in each category is presented in [Table 2](#).

The geographic distribution of studies based on first author’s affiliated institution revealed that the majority of research originated from North America (n=42, 48.3%), followed by Asia (n=26, 29.9%) and Europe (n=18, 20.7%). Oceania was represented by a single study (1.1%) from Australia, while no research related to LLMs in ophthalmology was reported from Latin America and the Caribbean or Africa. The distribution of studies by region is presented in [Table 3](#) and [Figure 1](#).

The majority of studies were published in high-impact journals, with 65 studies (74.7%) in Q1 journals, 10 studies (11.5%) in Q2 journals, and 9 studies (10.3%) in Q3 journals. In addition, 4 studies (4.6%) were not assigned a quartile, of which 3 were pre-prints in clinical decision-making. Most studies were published in 2024 (n=71, 81.6%), followed by 15 studies (17.2%) in 2023 and 1 study (1.1%) in 2022.

Gender disparity was evident for all author roles, with women accounting for only 29.9% of first authors, 25.3% of last authors, and 26.4% of corresponding authors ([Figure 2](#)). As expected, last and corresponding authors generally had greater scholastic impact than first authors. The average h-indices for last and corresponding authors were 39.7±37.7 and 28.7±33.7, respectively, compared to 8.9±6.2 for first authors (p<0.001). Similarly, the average i10-indices for last and corresponding authors were 147.1±254.2 and 94.4±222.2, compared to 12.0±14.9 for first authors (p=0.005) ([Table 4](#)).

We also evaluated the publications according to each application category, and the results of this in-depth analysis are as follows:

Clinical Decision-Making

Studies within this category were published from a diverse range of countries across several continents, although the majority of them still originated from the USA (n=17, 44.7%). Most studies were published within the past year in 2024 (n=32, 84.2%). Gender disparities were evident across all subgroups, most notably in the cornea/

anterior segment subspecialty, where there were no women as first, last, or corresponding authors ([Table 5](#)).

Most studies were published in high-impact journals, particularly within the retina, glaucoma, and multiple subspecialties subgroups, with 80.0%, 100%, and 87.5% appearing in Q1 journals, respectively. The cornea/anterior segment and uveitis subgroups had a mix of Q1, Q2, and/or Q3 journals. Notably, no studies were published in Q4 journals. A detailed list of the studies in this category is provided in the supplementary material ([Supplementary Table 1](#)).¹⁷⁻⁵⁴

Table 2. Distribution of studies within the clinical decision-making category (n=38)

Clinical decision-making	Number of studies	Percentage
Retina	15	39.5%
Glaucoma	6	15.8%
Cornea/anterior segment	3	7.9%
Uveitis	3	7.9%
Neuro-ophthalmology	2	5.3%
Pediatrics	1	2.6%
Multiple subspecialties	8	21.1%

Table 3. Geographical distribution of studies according to first author’s affiliated institution (n=87)

Region/continent	Country	Number of studies	Percentage
North America	United States	38	43.7%
	Canada	4	4.6%
Europe	United Kingdom	9	10.3%
	Germany	3	3.5%
	Switzerland	2	2.3%
	Italy	2	2.3%
	Romania	1	1.1%
	Finland	1	1.1%
Asia	China	12	13.8%
	Singapore	4	4.6%
	Türkiye	4	4.6%
	India	3	3.5%
	Japan	1	1.1%
	Israel	1	1.1%
Oceania	Jordan	1	1.1%
	Australia	1	1.1%
Africa		0	0%
Latin America and the Caribbean		0	0%

Table 1. Distribution of studies according to application types (n=87)

Category	Number of studies	Percentage
Clinical decision-making	38	43.7%
Educational applications	22	25.3%
Patient interaction applications	18	20.7%
Miscellaneous applications	9	10.3%

Educational Applications

Of the 22 studies, North America contributed the largest share (45.5%), with 8 studies from the USA and 2 from Canada. Europe accounted for 31.8% of the studies, with 6 studies from the UK and 1 study from Germany. Asia contributed 5 studies (22.7%), with 2 from China and 1 each from Türkiye, Israel, and Japan. Most of these studies were published in Q1 journals (72.7%), followed by Q3 journals (18.2%) and Q2 journals (9.1%), with the majority published in 2024 (n=18, 81.8%). Gender disparity was also evident in this group, with male dominance in all author roles. Women accounted for 4 first authors (18.2%), 1 last author (4.5%), and 3 corresponding authors (13.6%).

Patient Interaction Applications

Studies in this group were primarily from Asia (44.4%), followed by North America (38.9%), Europe (11.1%), and Oceania (5.6%). The majority of the studies were published in Q1 journals (66.7%), with smaller portions in Q2 and Q3 journals. Female authors were overall more represented in this category, accounting for 33.3% of first authors, 5.6% of last authors, and 44.4% of corresponding authors.

Miscellaneous Applications

Of the 9 studies, North America made the largest contribution, accounting for 66.7% of the studies, with

6 studies originating from the USA. Türkiye, India, and Finland contributed the remaining studies, each providing one. The distribution of author roles was nearly balanced for first and corresponding authors, with 44.4% of first authors being men and 55.6% women, while 55.6% of corresponding authors were men and 44.4% were women. However, last authors were still predominantly male (66.7%).

The distribution of gender and geographical regions for first and last authors across educational applications, patient interaction applications, and miscellaneous applications is detailed in [Table 6](#). A comprehensive list of studies in the other categories and overall journal metrics are provided in the supplementary material ([Supplementary Tables 2 and 3](#)).⁵⁵⁻¹⁰³

Discussion

LLMs are generative AI systems that perform natural language processing in order to understand human text and speech.^{104,105} ChatGPT, BERT, and LLaMA are commonly used LLMs that employ deep learning (DL) techniques to engage in meaningful conversations with users.¹⁰⁴ These applications provide access to vast amounts of knowledge and may also help patients access medical information relevant to their specific conditions, assess the urgency of their symptoms, or be directed to the appropriate subspecialty.¹⁰⁶

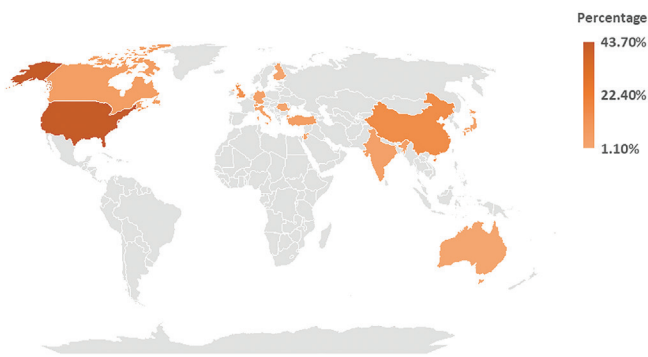


Figure 1. Heat map distribution of global LLM-related studies in ophthalmology
LLM: Large language model

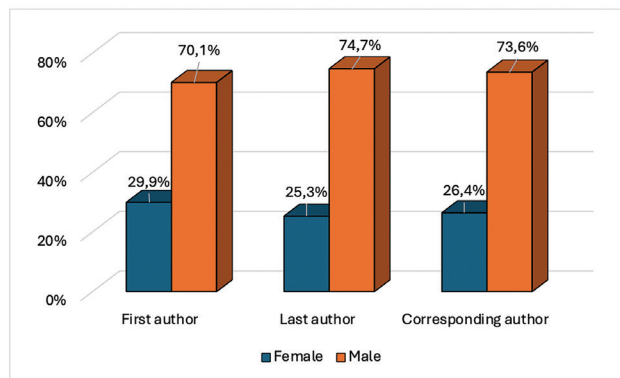


Figure 2. Gender distribution among authorship roles (presented as percentage)

Table 4. Distribution of genders and scholastic impact factors (h-index and i10-index) across first, last, and corresponding author roles

Author role	Female	Male	h-index (Mean ± SD)	i10-index (Mean ± SD)
First author	26 (29.9%)	61 (70.1%)	8.9±6.2	12.0±14.9
Last author	22 (25.3%)	65 (74.7%)	39.7±37.7	147.1±254.2
Corresponding author	23 (26.4%)	64 (73.6%)	28.7±33.7	94.4±222.2

SD: Standard deviation

Table 5. Detailed distribution of gender and geographical region of studies according to the first and last authors' affiliated institutions for clinical decision-making applications

	First author			Last author		
	Gender	Geographical region	Top 3 countries of publication	Gender	Geographical region	Top 3 countries of publication
Retina ^{17,31} (n=15)	Female: 4 (26.7%) Male: 11 (73.3%)	Asia: 6 (40.0%) Europe: 5 (33.3%) North America: 4 (26.7%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	China: 5 (33.3%) UK: 3 (20.0%) USA: 3 (20.0%)	Female: 3 (20.0%) Male: 12 (80.0%)	Asia: 6 (40.0%) Europe: 5 (33.3%) North America: 4 (26.7%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	China: 5 (33.3%) UK: 4 (26.7%) USA: 3 (20.0%)
Glaucoma ^{32,33,34,35,36,37} (n=6)	Female: 2 (33.3%) Male: 4 (66.7%)	North America: 4 (66.7%) Asia: 1 (16.7%) Europe: 1 (16.7%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 4 (66.7%) China: 1 (16.7%) Italy: 1 (16.7%)	Female: 0 (0%) Male: 6 (100%)	North America: 4 (66.7%) Asia: 1 (16.7%) Europe: 1 (16.7%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 4 (66.7%) China: 1 (16.7%) Italy: 1 (16.7%)
Cornea/anterior segment ^{38,39,40} (n=3)	Female: 0 (0%) Male: 3 (100%)	North America: 2 (66.7%) Europe: 1 (33.3%) Asia: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 2 (66.7%) Germany: 1 (33.3%)	Female: 0 (0%) Male: 3 (100%)	North America: 2 (66.7%) Europe: 1 (33.3%) Asia: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 2 (66.7%) Germany: 1 (33.3%)
Uveitis ^{41,42,43} (n=3)	Female: 2 (66.7%) Male: 1 (33.3%)	Asia: 1 (33.3%) Europe: 1 (33.3%) North America: 1 (33.3%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	Singapore: 1 (33.3%) Switzerland: 1 (33.3%) USA: 1 (33.3%)	Female: 2 (66.7%) Male: 1 (33.3%)	Asia: 1 (33.3%) Europe: 1 (33.3%) North America: 1 (33.3%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	Singapore: 1 (33.3%) UK: 1 (33.3%) USA: 1 (33.3%)
Neuro-ophthalmology ^{44,45} (n=2)	Female: 1 (50%) Male: 1 (50%)	North America: 2 (100%) Asia: 0 (0%) Europe: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 2 (100%)	Female: 0 (0%) Male: 2 (100%)	North America: 2 (100%) Asia: 0 (0%) Europe: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 2 (100%)

	First author			Last author		
	Gender	Geographical region	Top 3 countries of publication	Gender	Geographical region	Top 3 countries of publication
	Pediatrics ⁴⁶ (n=1)	Female: 0 (0%) Male: 1 (100%)	North America: 1 (100%) Asia: 0 (0%) Europe: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 1 (100%)	Female: 0 (0%) Male: 1 (100%)	North America: 1 (100%) Asia: 0 (0%) Europe: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)
Multiple subspecialties ^{47,48,49,50,51,52,53,54} (n=8)	Female: 1 (12.5%) Male: 7 (87.5%)	North America: 5 (62.5%) Asia: 3 (37.5%) Europe: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 4 (50%) China: 3 (37.5%) Canada: 1 (12.5%)	Female: 3 (37.5%) Male: 5 (62.5%)	North America: 5 (62.5%) Asia: 3 (37.5%) Europe: 0 (0%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 4 (50.0%) China: 3 (37.5%) Canada: 1 (12.5%)

As AI applications continue to grow in medicine, there has also been a rising interest in their use within ophthalmology. Over the past year, ophthalmology has seen an increase in the diverse applications of LLMs, ranging from clinical decision-making to enhancing patient interactions. The purpose of our study was to explore the current state of LLM research in ophthalmology through a bibliographic analysis.

We observed that most of the LLM studies in ophthalmology focused on clinical decision-making, with the majority aimed at retinal applications. This interest in clinical decision-making applications is not surprising, as LLMs have shown promise in assisting with complex diagnostic and treatment decisions. However, there is still significant uncertainty regarding how LLMs could be integrated into actual patient care from a regulatory standpoint.¹⁰⁷ While the reasons for the high number of retina-related LLM publications are not entirely clear, we speculate that it may be due to the history of DL applications in ophthalmology. DL-based image analyses in ophthalmology have been spearheaded by the retina subspecialty. Therefore, the same group of researchers who focused on DL-based analysis of retinal images may have been early adopters of LLM-related research.¹⁰⁸

Similarly, our research team has recently focused on LLMs. For example, we published a meta-analysis evaluating the accuracy of LLMs in answering board-style questions.⁵⁵ In addition, we further examined the impact of integrating retrieval-augmented generation in enhancing LLM performance on both text-based and image-based board-style questions.^{109,110} Another area of interest has been the difference in performance between different LLMs in answering questions related to social determinants of health in ophthalmology.¹¹¹

Furthermore, our current study also demonstrates that LLMs possess significant scientific merit, as evidenced by their increasing presence in top-tier publications in recent years. These studies present substantial scholarly impact, with a notable number appearing in top-tier Q1 journals. We observed that glaucoma-related applications exhibited the greatest scientific impact in clinical decision-making, as reflected by their high h-index and i10-index values, along with their strong presence in Q1 journals. However, overall, the most scientifically impactful applications were those involving algorithms trained for educational purposes, such as evaluating their performance in clinical knowledge exams compared to professionals or developing educational materials. These studies demonstrated higher average h-index and i10-index values, with the majority published in Q1 journals. The high impact of educational applications in particular may be attributed to their adaptability across medical disciplines beyond ophthalmology, their potential

Table 6. Detailed distribution of gender and geographical region of studies according to the first and last authors' affiliated institutions for educational, patient interaction, and miscellaneous applications

	First author			Last author		
	Gender	Geographical region	Top 3 countries of publication	Gender	Geographical region	Top 3 countries of publication
Educational applications ⁵⁵⁻⁷⁶	Female: 4 (18.2%) Male: 18 (81.8%)	North America: 10 (45.5%) Europe: 7 (31.8%) Asia: 5 (22.7%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 8 (36.4%) UK: 6 (27.3%) Canada: 2 (9.1%) China: 2 (9.1%)	Female: 1 (4.5%) Male: 21 (95.5%)	North America: 12 (54.5%) Europe: 5 (22.7%) Asia: 5 (22.7%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 9 (40.9%) Canada: 3 (13.6%) China: 2 (9.1%)
Patient interaction applications ⁷⁷⁻⁹⁴	Female: 6 (33.3%) Male: 12 (66.7%)	Asia: 8 (44.4%) North America: 7 (38.9%) Europe: 2 (11.1%) Oceania: 1 (5.6%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%)	USA: 7 (38.9%) Singapore: 3 (16.7%) Türkiye: 2 (11.1%)	Female: 8 (55.6%) Male: 10 (44.4%)	Asia: 8 (44.4%) North America: 7 (38.9%) Europe: 2 (11.1%) Oceania: 1 (5.6%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%)	USA: 7 (38.9%) Singapore: 3 (16.7%) Türkiye: 2 (11.1%)
Miscellaneous applications ^{95-96,97,98,99,100,101,102,103}	Female: 5 (55.6%) Male: 4 (44.4%)	North America: 6 (66.7%) Asia: 2 (22.2%) Europe: 1 (11.1%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 6 (66.7%) India: 1 (11.1%) Türkiye: 1 (11.1%)	Female: 3 (33.3%) Male: 6 (66.7%)	North America: 6 (66.7%) Asia: 2 (22.2%) Europe: 1 (11.1%) Africa: 0 (0%) Latin America/Caribbean: 0 (0%) Oceania: 0 (0%)	USA: 6 (66.7%) UK: 1 (11.1%) Türkiye: 1 (11.1%)

to enhance training outcomes, and their effectiveness in addressing educational gaps in medical training.

Despite the increasing prominence of LLMs, the underrepresentation of female authors remains an ongoing challenge. A recent study showed an increase in the proportion of women in research since 2018, from 33% to 37% of first authors and 27% to 30% of last authors.^{112,113} However, gender disparity continues to be a significant issue, particularly in the field of AI. Similar to the literature, our study revealed a prominent gender gap in all author roles, with women accounting for only 29.9% of first authors, 25.3% of last authors, and 26.4% of corresponding authors. This gender distribution in our study suggests that despite recent efforts to increase female representation in academic publishing, significant barriers still exist for women in reaching key author roles. This underscores the need for continued efforts to promote gender equity. The notable underrepresentation of female authors likely reflects broader systemic issues, such as gender disparities in science, technology, engineering and mathematics education and leadership roles, limited mentorship opportunities for women in AI-driven medical research, and potential institutional biases.¹¹⁴

A similar disparity can be observed in the geographical distribution of studies. Research has shown that English-speaking countries, particularly the US (48.2%), dominate authorship in leading medical journals. In contrast, authors from developing countries remain underrepresented, although there has been an increase in geographical diversity in recent years.¹¹⁵ Another recent study also highlighted a similar bias in the rejection of research submitted to journals. Compared to authors from institutions in non-Western countries, those from Western countries are 5.7% more likely to have their manuscript accepted after rejection. Additionally, authors from Western countries tend to publish 23 days faster, revise 5.9% less often, change co-authors 12.0% less frequently, and ultimately publish in journals with impact factors that are 0.8% higher.^{11,116}

Consistent with the literature, our study revealed a similar disparity in the geographical origin of research based on first authorship. Overall, English-speaking countries dominated the research output, with the USA contributing 43.7% of the studies and the UK contributing to 10.3%. However, China ranked a strong second after the USA, contributing 13.8% of the studies, underscoring its significant contribution to the field. This suggests that while the USA and China are leading research in LLMs in ophthalmology, there is still a strong need for contributions from other regions, particularly in the development of applications tailored to the specific cultural and societal needs of those regions. This imbalance may be partly explained by the predominance of English-language data used for training LLMs.¹¹⁷ Furthermore, it may also be attributed to the immense investment into fundamental AI infrastructure in the US and China, funding availability, and access to advanced technologies, all of which play a critical role in shaping the global distribution of LLM-related research.¹¹⁸

Study Limitations

This study also has limitations, particularly the use of PubMed as the sole platform for the literature search. While other databases such as arXiv and IEEE Xplore host a significant number of AI-related studies, we chose to include only those indexed in PubMed due to its high standards for peer-reviewed content, thereby ensuring the scientific rigor of the included articles. Additionally, we recognize that categorizing studies based on AI-related infrastructure or funding opportunities could offer a different perspective on the distribution of LLM-related research and should be explored in future studies.

Conclusion

The use of LLMs in ophthalmology is rapidly gaining interest, with the majority of studies published recently (in 2024) and in top-tier (Q1) journals. North America leads in publications, followed by growing contributions from Asia and Europe, while other regions, including Oceania, Latin America and the Caribbean, and Africa, remain underrepresented. A similar imbalance is observed in the gender distribution in authorship, with women being severely underrepresented across all key author roles. These geographic and gender inequities highlight significant gaps in global and demographic representation within LLM research in ophthalmology and demonstrate the need for further progress.

Declarations

Authorship Contributions

Concept: N.D.K., T.Y.A.L., Design: N.D.K., T.Y.A.L., Data Collection or Processing: N.D.K., T.Y.A.L., Analysis or Interpretation: N.D.K., T.Y.A.L., Literature Search: N.D.K., T.Y.A.L., Writing: N.D.K., T.Y.A.L.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

References

1. Amisha, Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. 2019;8:2328-2331.
2. Ray P. ChatGPT: a comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*. 2023;3:121-154.
3. Large Language Model Introducing LLaMA 3.1: Our most capable models to date, 2024. <https://ai.meta.com/blog/meta-llama-3-1/>
4. Kocoń J, Cichecki I, Kaszyca O, Kochanek M, Szydło D, Baran J, Bielaniec J, Gruza M, Janz A, Kanclerz K, Kocoń A, Koptyra B, Mielezyczenko-Kowszewicz W, Miłkowski P, Oleksy M, Piasecki M, Radliński Ł, Wojtasik K, Woźniak S, Kazienko P. ChatGPT: Jack of all trades, master of none. *Information Fusion*. 2023;99:101861.
5. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. 2018.
6. Alowais SA, Alghamdi SS, Alsuhbany N, Alqahtani T, Alshaya AI, Almohareb SN, Aldairem A, Alrashed M, Bin Saleh K, Badreldin HA, Al Yami MS, Al Harbi S, Albekairy AM. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023;23:689.
7. Nakaura T, Ito R, Ueda D, Nozaki T, Fushimi Y, Matsui Y, Yanagawa M, Yamada A, Tsuboyama T, Fujima N, Tatsugami F, Hirata K, Fujita S, Kamagata K, Fujioka T, Kawamura M, Naganawa S. The impact of large language models on radiology: a guide for radiologists on the latest innovations in AI. *Jpn J Radiol*. 2024;42:685-696.
8. Wyatt KD, Alexander N, Hills GD, Liang WH, Kadauke S, Volchenboum SL, Mian A, Phillips CA. Making sense of artificial intelligence and large language models-including ChatGPT-in pediatric hematology/oncology. *Pediatr Blood Cancer*. 2024;71:e31143.
9. Quer G, Topol EJ. The potential for large language models to transform cardiovascular medicine. *Lancet Digit Health*. 2024;6:e767-e771.
10. Tan TF, Thirunavukarasu AJ, Campbell JP, Keane PA, Pasquale LR, Abramoff MD, Kalpathy-Cramer J, Lum F, Kim JE, Baxter SL, Ting DSW. Generative artificial intelligence through ChatGPT and other large language models in ophthalmology: clinical applications and challenges. *Ophthalmol Sci*. 2023;3:100394.
11. Draux, H. Research on artificial intelligence – the global divides. January 4, 2024. <https://www.digital-science.com/blog/research-on-artificial-intelligence-the-global-divides/>
12. Stanford Institute for Human-Centered AI (HAI). Artificial Intelligence Index Report 2024. <https://aiindex.stanford.edu/report/>
13. Pal S, Lazzaroni RM, Mendoza P. AI's missing link: the gender gap in the talent pool. October 10, 2024. <https://www.stiftung-nv.de/publications/ai-gender-gap/>

14. United Nations, D. o. E. a. S. A., Population Division World Population Prospects 2024. <https://population.un.org/wpp/definition-of-regions>
15. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A*. 2005;102:16569-16572.
16. Cornell University Library. Measuring your research impact: i10-Index. <https://guides.library.cornell.edu/c.php?g=32272&p=203393>
17. Ferro Desideri L, Roth J, Zinkernagel M, Anguita R. Application and accuracy of artificial intelligence-derived large language models in patients with age related macular degeneration. *Int J Retina Vitreous*. 2023;9:71.
18. Li J, Guan Z, Wang J, Cheung CY, Zheng Y, Lim LL, Lim CC, Ruamviboonsuk P, Raman R, Corsino L, Echouffo-Tcheugui JB, Luk AOY, Chen LJ, Sun X, Hamzah H, Wu Q, Wang X, Liu R, Wang YX, Chen T, Zhang X, Yang X, Yin J, Wan J, Du W, Quek TC, Goh JHL, Yang D, Hu X, Nguyen TX, Szeto SKH, Chotcomwongse P, Malek R, Normatova N, Ibragimova N, Srinivasan R, Zhong P, Huang W, Deng C, Ruan L, Zhang C, Zhang C, Zhou Y, Wu C, Dai R, Koh SWC, Abdullah A, Hee NKY, Tan HC, Liew ZH, Tien CS, Kao SL, Lim AYL, Mok SF, Sun L, Gu J, Wu L, Li T, Cheng D, Wang Z, Qin Y, Dai L, Meng Z, Shu J, Lu Y, Jiang N, Hu T, Huang S, Huang G, Yu S, Liu D, Ma W, Guo M, Guan X, Yang X, Bascaran C, Cleland CR, Bao Y, Ekinci EI, Jenkins A, Chan JCN, Bee YM, Sivaprasad S, Shaw JE, Simó R, Keane PA, Cheng CY, Tan GSW, Jia W, Tham YC, Li H, Sheng B, Wong TY. Integrated image-based deep learning and language models for primary diabetes care. *Nat Med*. 2024;30:2886-2896.
19. Anguita R, Makuloluwa A, Hind J, Wickham L. Large language models in vitreoretinal surgery. *Eye (Lond)*. 2024;38:809-810.
20. Anguita R, Downie C, Ferro Desideri L, Sagoo MS. Assessing large language models' accuracy in providing patient support for choroidal melanoma. *Eye (Lond)*. 2024;38:3113-3117.
21. 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.
22. Chen X, Zhang W, Xu P, Zhao Z, Zheng Y, Shi D, He M. FFA-GPT: an automated pipeline for fundus fluorescein angiography interpretation and question-answer. *NPJ Digit Med*. 2024;7:111.
23. Carlà MM, Gambini G, Baldascino A, Giannuzzi F, Boselli F, Crincoli E, D'Onofrio NC, Rizzo S. Exploring AI-chatbots' capability to suggest surgical planning in ophthalmology: ChatGPT versus Google Gemini analysis of retinal detachment cases. *Br J Ophthalmol*. 2024;108:1457-1469.
24. 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.
25. Ghalibafan S, Taylor Gonzalez DJ, Cai LZ, Graham Chou B, Panneerselvam S, Conrad Barrett S, Djulbegovic MB, Yannuzzi NA. Applications of multimodal generative artificial intelligence in a real-world retina clinic setting. *Retina*. 2024;44:1732-1740.
26. Chen X, Xu P, Li Y, Zhang W, Song F, He M, Shi D. ChatFFA: an ophthalmic chat system for unified vision-language understanding and question answering for fundus fluorescein angiography. *iScience*. 2024;27:110021.
27. Chen X, Zhang W, Zhao Z, Xu P, Zheng Y, Shi D, He M. ICGA-GPT: report generation and question answering for indocyanine green angiography images. *Br J Ophthalmol*. 2024;108:1450-1456.
28. Gopalakrishnan N, Joshi A, Chhablani J, Yadav NK, Reddy NG, Rani PK, Pulipaka RS, Shetty R, Sinha S, Prabhu V, Venkatesh R. Recommendations for initial diabetic retinopathy screening of diabetic patients using large language model-based artificial intelligence in real-life case scenarios. *Int J Retina Vitreous*. 2024;10:11.
29. Balas M, Mandelcorn ED, Yan P, Ing EB, Crawford SA, Arjmand P. ChatGPT and retinal disease: a cross-sectional study on AI comprehension of clinical guidelines. *Can J Ophthalmol*. 2024;60:e117-e123.
30. Liu X, Wu J, Shao A, Shen W, Ye P, Wang Y, Ye J, Jin K, Yang J. Uncovering language disparity of ChatGPT on retinal vascular disease classification: cross-sectional study. *J Med Internet Res*. 2024;26:e51926.
31. Tailor PD, Dalvin LA, Chen JJ, Iezzi R, Olsen TW, Scruggs BA, Barkmeier AJ, Bakri SJ, Ryan EH, Tang PH, Parke DW 3rd, Belin PJ, Sridhar J, Xu D, Kuriyan AE, Yonekawa Y, Starr MR. 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.
32. Carlà MM, Gambini G, Baldascino A, Boselli F, Giannuzzi F, Margollicci F, Rizzo S. Large language models as assistance for glaucoma surgical cases: a ChatGPT vs. Google Gemini comparison. *Graefes Arch Clin Exp Ophthalmol*. 2024;262:2945-2959.
33. Delsoz M, Raja H, Madadi Y, Tang AA, Wiroszko BM, Kahook MY, Yousefi S. The use of ChatGPT to assist in diagnosing glaucoma based on clinical case reports. *Ophthalmol Ther*. 2023;12:3121-3132.
34. Huang X, Raja H, Madadi Y, Delsoz M, Poursorouh A, Kahook MY, Yousefi S. Predicting glaucoma before onset using a large language model chatbot. *Am J Ophthalmol*. 2024;266:289-299.
35. Xue X, Zhang D, Sun C, Shi Y, Wang R, Tan T, Gao P, Fan S, Zhai G, Hu M, Wu Y. Xiaoqing: a Q&A model for glaucoma based on LLMs. *Comput Biol Med*. 2024;174:108399.
36. Raja H, Huang X, Delsoz M, Madadi Y, Poursorouh A, Munawar A, Kahook MY, Yousefi S. Diagnosing glaucoma based on the ocular hypertension treatment study dataset using chat generative pre-trained transformer as a large language model. *Ophthalmol Sci*. 2025;5:100599.
37. AlRyalat SA, Musleh AM, Kahook MY. Evaluating the strengths and limitations of multimodal ChatGPT-4 in detecting glaucoma using fundus images. *Front Ophthalmol (Lausanne)*. 2024;4:1387190.
38. Ćirković A, Katz T. Exploring the Potential of ChatGPT-4 in predicting refractive surgery categorizations: comparative study. *JMIR Form Res*. 2023;7:e51798.
39. Delsoz M, Madadi Y, Raja H, Munir WM, Tamm B, Mehravaran S, Soleimani M, Djalilian A, Yousefi S. Performance of ChatGPT in diagnosis of corneal eye diseases. *Cornea*. 2024;43:664-670.
40. Tuttle JJ, Moshirfar M, Garcia J, Altaf AW, Omidvarnia S, Hoopes PC. Learning the Randleman criteria in refractive surgery: utilizing ChatGPT-3.5 versus internet search engine. *Cureus*. 2024;16:e64768.
41. Rojas-Carabali W, Sen A, Agarwal A, Tan G, Cheung CY, Rousselot A, Agrawal R, Liu R, Cifuentes-González C, Elze T, Kempen JH, Sobrin L, Nguyen QD, de-la-Torre A, Lee B, Gupta V, Agrawal R. Chatbots Vs. human experts: evaluating diagnostic performance of chatbots in uveitis and the perspectives on AI adoption in ophthalmology. *Ocul Immunol Inflamm*. 2024;32:1591-1598.
42. Schumacher I, Bühler VMM, Jaggi D, Roth J. Artificial intelligence derived large language model in decision-making process in uveitis. *Int J Retina Vitreous*. 2024;10:63.
43. Marshall RF, Mallem K, Xu H, Thorne J, Burkholder B, Chaon B, Liberman P, Berkenstock M. Investigating the accuracy and completeness of an artificial intelligence large language model about uveitis: an evaluation of ChatGPT. *Ocul Immunol Inflamm*. 2024;32:2052-2055.

44. Madadi Y, Delsoz M, Lao PA, Fong JW, Hollingsworth TJ, Kahook MY, Yousefi S. ChatGPT assisting diagnosis of neuro-ophthalmology diseases based on case reports. medRxiv [Preprint]. 2023:2023.
45. Tailor PD, Dalvin LA, Starr MR, Tajfirouz DA, Chodnicki KD, Brodsky MC, Mansukhani SA, Moss HE, Lai KE, Ko MW, Mackay DD, Di Nome MA, Dumitrascu OM, Pless ML, Eggenberger ER, Chen JJ. A comparative study of large language models, human experts, and expert-edited large language models to neuro-ophthalmology questions. *J Neuroophthalmol*. 2025;45:71-77.
46. Upadhyaya DP, Shaikh AG, Cakir GB, Prantzas K, Golnari P, Ghasia FF, Sahoo SS. A 360° view for large language models: early detection of amblyopia in children using multi-view eye movement recordings. medRxiv [Preprint]. 2024:2024.
47. Luo MJ, Pang J, Bi S, Lai Y, Zhao J, Shang Y, Cui T, Yang Y, Lin Z, Zhao L, Wu X, Lin D, Chen J, Lin H. Development and evaluation of a retrieval-augmented large language model framework for ophthalmology. *JAMA Ophthalmol*. 2024;142:798-805.
48. Zheng C, Ye H, Guo J, Yang J, Fei P, Yuan Y, Huang D, Huang Y, Peng J, Xie X, Xie M, Zhao P, Chen L, Zhang M. Development and evaluation of a large language model of ophthalmology in Chinese. *Br J Ophthalmol*. 2024;108:1390-1397.
49. Huang AS, Hirabayashi K, Barna L, Parikh D, Pasquale LR. Assessment of a large language model's responses to questions and cases about glaucoma and retina management. *JAMA Ophthalmol*. 2024;142:371-375.
50. Deng Z, Gao W, Chen C, Niu Z, Gong Z, Zhang R, Cao Z, Li F, Ma Z, Wei W, Ma L. OphGLM: an ophthalmology large language-and-vision assistant. *Artif Intell Med*. 2024;157:103001.
51. Haghghi T, Gholami S, Sokol JT, Kishnani E, Ahsaniyan A, Rahmanian H, Hedayati F, Leng T, Alam MN. EYE-LLaMA, an in-domain large language model for ophthalmology. bioRxiv [Preprint]. 2025.
52. Chen JS, Reddy AJ, Al-Sharif E, Shoji MK, Kalaw FGP, Eslani M, Lang PZ, Arya M, Koretz ZA, Bolo KA, Arnett JJ, Roginiel AC, Do JL, Robbins SL, Camp AS, Scott NL, Rudell JC, Weinreb RN, Baxter SL, Granet DB. Analysis of ChatGPT responses to ophthalmic cases: can ChatGPT think like an ophthalmologist? *Ophthalmol Sci*. 2024;5:100600.
53. Milad D, Antaki F, Milad J, Farah A, Khairy T, Mikhail D, Giguère CÉ, Touma S, Bernstein A, Szigiato AA, Nayman T, Mullie GA, Duval R. Assessing the medical reasoning skills of GPT-4 in complex ophthalmology cases. *Br J Ophthalmol*. 2024;108:1398-1405.
54. Hu X, Ran AR, Nguyen TX, Szeto S, Yam JC, Chan CKM, Cheung CY. What can GPT-4 do for diagnosing rare eye diseases? A pilot study. *Ophthalmol Ther*. 2023;12:3395-3402.
55. Wu JH, Nishida T, Liu TYA. Accuracy of large language models in answering ophthalmology board-style questions: a meta-analysis. *Asia Pac J Ophthalmol (Phila)*. 2024;13:100106.
56. Raimondi R, Tzoumas N, Salisbury T, Di Simplicio S, Romano MR; North East Trainee Research in Ophthalmology Network (NETriON). Comparative analysis of large language models in the Royal College of Ophthalmologists fellowship exams. *Eye (Lond)*. 2023;37:3530-3533.
57. Sakai D, Maeda T, Ozaki A, Kanda GN, Kurimoto Y, Takahashi M. Performance of ChatGPT in board examinations for specialists in the Japanese Ophthalmology Society. *Cureus*. 2023;15:e49903.
58. Fowler T, Pullen S, Birkett L. Performance of ChatGPT and Bard on the official part 1 FRCOphth practice questions. *Br J Ophthalmol*. 2024;108:1379-1383.
59. Yaïci R, Cieplucha M, Bock R, Moayed F, Bechrakis NE, Berens P, Feltgen N, Friedburg D, Gräf M, Guthoff R, Hoffmann EM, Hoerauf H, Hintschich C, Kohnen T, Messmer EM, Nentwich MM, Pleyer U, Schaudig U, Seitz B, Geerling G, Roth M. ChatGPT und die deutsche Facharztprüfung für Augenheilkunde: eine Evaluierung [ChatGPT and the German board examination for ophthalmology: an evaluation]. *Ophthalmologie*. 2024;121:554-564.
60. Ming S, Guo Q, Cheng W, Lei B. Influence of model evolution and system roles on ChatGPT's performance in Chinese medical licensing exams: comparative study. *JMIR Med Educ*. 2024;10:e52784.
61. Bahir D, Zur O, Attal L, Nujeidat Z, Knaanie A, Pikkell J, Mimouni M, Plopsky G. Gemini AI vs. ChatGPT: a comprehensive examination alongside ophthalmology residents in medical knowledge. *Graefes Arch Clin Exp Ophthalmol*. 2025;263:527-536.
62. Gill GS, Tsai J, Moxam J, Sanghvi HA, Gupta S. Comparison of Gemini advanced and ChatGPT 4.0's performances on the ophthalmology resident ophthalmic knowledge assessment program (OKAP) examination review question banks. *Cureus*. 2024;16:e69612.
63. Antaki F, Milad D, Chia MA, Giguère CÉ, Touma S, El-Khoury J, Keane PA, Duval R. Capabilities of GPT-4 in ophthalmology: an analysis of model entropy and progress towards human-level medical question answering. *Br J Ophthalmol*. 2024;108:1371-1378.
64. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in ophthalmology: an analysis of its successes and shortcomings. *Ophthalmol Sci*. 2023;3:100324.
65. 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.
66. Dihan Q, Chauhan MZ, Eleiwa TK, Brown AD, Hassan AK, Khodeiry MM, Elsheikh RH, Oke I, Nihalani BR, VanderVeen DK, Sallam AB, Elhusseiny AM. Large language models: a new frontier in paediatric cataract patient education. *Br J Ophthalmol*. 2024;108:1470-1476.
67. Dihan Q, Chauhan MZ, Eleiwa TK, Hassan AK, Sallam AB, Khouri AS, Chang TC, Elhusseiny AM. Using large language models to generate educational materials on childhood glaucoma. *Am J Ophthalmol*. 2024;265:28-38.
68. Dihan QA, Brown AD, Zaldivar AT, Chauhan MZ, Eleiwa TK, Hassan AK, Solyman O, Gise R, Phillips PH, Sallam AB, Elhusseiny AM. Advancing patient education in idiopathic intracranial hypertension: the promise of large language models. *Neurol Clin Pract*. 2025;15:e200366.
69. Singer MB, Fu JJ, Chow J, Teng CC. Development and evaluation of aeyeconsult: a novel ophthalmology chatbot leveraging verified textbook knowledge and GPT-4. *J Surg Educ*. 2024;81:438-443.
70. Waisberg E, Ong J, Masalkhi M, Lee AG. Large language model (LLM)-driven chatbots for neuro-ophthalmic medical education. *Eye (Lond)*. 2024;38:639-641.
71. Durmaz Engin C, Karatas E, Ozturk T. Exploring the Role of ChatGPT-4, BingAI, and Gemini as virtual consultants to educate families about retinopathy of prematurity. *Children (Basel)*. 2024;11:750.
72. Jung H, Oh J, Stephenson KAJ, Joe AW, Mammo ZN. Prompt engineering with ChatGPT3.5 and GPT4 to improve patient education on retinal diseases. *Can J Ophthalmol*. 2025;60:e375-e381.
73. Cai LZ, Shaheen A, Jin A, Fukui R, Yi JS, Yannuzzi N, Alabiad C. Performance of generative large language models on ophthalmology board-style questions. *Am J Ophthalmol*. 2023;254:141-149.
74. Thirunavukarasu AJ, Mahmood S, Malem A, Foster WP, Sanghera R, Hassan R, Zhou S, Wong SW, Wong YL, Chong YJ, Shakeel A, Chang YH, Tan BKJ, Jain N, Tan TE, Rauz S, Ting DSW, Ting DSJ. Large language models approach expert-level clinical knowledge and reasoning in ophthalmology: a head-to-head cross-sectional study. *PLOS Digit Health*. 2024;3:e0000341.

75. 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.
76. Chen X, Zhao Z, Zhang W, Xu P, Wu Y, Xu M, Gao L, Li Y, Shang X, Shi D, He M. EyeGPT for patient inquiries and medical education: development and validation of an ophthalmology large language model. *J Med Internet Res*. 2024;26:e60063.
77. Baxter SL, Longhurst CA, Millen M, Sitapati AM, Tai-Seale M. Generative artificial intelligence responses to patient messages in the electronic health record: early lessons learned. *JAMIA Open*. 2024;7:ooae028.
78. Cohen SA, Brant A, Fisher AC, Pershing S, Do D, Pan C. Dr. Google vs. Dr. ChatGPT: exploring the use of artificial intelligence in ophthalmology by comparing the accuracy, safety, and readability of responses to frequently asked patient questions regarding cataracts and cataract surgery. *Semin Ophthalmol*. 2024;39:472-479.
79. Muntean GA, Marginean A, Groza A, Damian I, Roman SA, Hapca MC, Sere AM, Mănoiu RM, Muntean MV, Nicoară SD. A qualitative evaluation of ChatGPT4 and PaLM2's response to patient's questions regarding age-related macular degeneration. *Diagnostics (Basel)*. 2024;14:1468.
80. Strzalkowski P, Strzalkowska A, Chhablani J, Pfau K, Errera MH, Roth M, Schaub F, Bechrakis NE, Hoerauf H, Reiter C, Schuster AK, Geerling G, Guthoff R. Evaluation of the accuracy and readability of ChatGPT-4 and Google Gemini in providing information on retinal detachment: a multicenter expert comparative study. *Int J Retina Vitreous*. 2024;10:61.
81. Kayabaşı M, Köksaldı S, Durmaz Engin C. Evaluating the reliability of the responses of large language models to keratoconus-related questions. *Clin Exp Optom*. 2025;108:784-791.
82. Shi R, Liu S, Xu X, Ye Z, Yang J, Le Q, Qiu J, Tian L, Wei A, Shan K, Zhao C, Sun X, Zhou X, Hong J. Benchmarking four large language models' performance of addressing Chinese patients' inquiries about dry eye disease: A two-phase study. *Heliyon*. 2024;10:e34391.
83. Tan DNH, Tham YC, Koh V, Loon SC, Aquino MC, Lun K, Cheng CY, Ngiam KY, Tan M. Evaluating Chatbot responses to patient questions in the field of glaucoma. *Front Med (Lausanne)*. 2024;11:1359073.
84. Ichhpujani P, Parmar UPS, Kumar S. Appropriateness and readability of Google Bard and ChatGPT-3.5 generated responses for surgical treatment of glaucoma. *Rom J Ophthalmol*. 2024;68:243-248.
85. Pushpanathan K, Lim ZW, Er Yew SM, Chen DZ, Hui'En Lin HA, Lin Goh JH, Wong WM, Wang X, Jin Tan MC, Chang Koh VT, Tham YC. Popular large language model chatbots' accuracy, comprehensiveness, and self-awareness in answering ocular symptom queries. *iScience*. 2023;26:108163.
86. Tailor PD, Xu TT, Fortes BH, Iezzi R, Olsen TW, Starr MR, Bakri SJ, Scruggs BA, Barkmeier AJ, Patel SV, Baratz KH, Bernhisel AA, Wagner LH, Tooley AA, Roddy GW, Sit AJ, Wu KY, Bothun ED, Mansukhani SA, Mohny BG, Chen JJ, Brodsky MC, Tajfirouz DA, Chodnicki KD, Smith WM, Dalvin LA. Appropriateness of ophthalmology recommendations from an online chat-based artificial intelligence model. *Mayo Clin Proc Digit Health*. 2024;2:119-128.
87. Alqudah AA, Aleshawi AJ, Baker M, Alnajjar Z, Ayasrah I, Ta'ani Y, Al Salkhadi M, Aljawarneh S. Evaluating accuracy and reproducibility of ChatGPT responses to patient-based questions in ophthalmology: an observational study. *Medicine (Baltimore)*. 2024;103:e39120.
88. Zandi R, Fahey JD, Drakopoulos M, Bryan JM, Dong S, Bryar PJ, Bidwell AE, Bowen RC, Lavine JA, Mirza RG. Exploring diagnostic precision and triage proficiency: a comparative study of GPT-4 and bard in addressing common ophthalmic complaints. *Bioengineering (Basel)*. 2024;11:120.
89. Bernstein IA, Zhang YV, Govil D, Majid I, Chang RT, Sun Y, Shue A, Chou JC, Schehlein E, Christopher KL, Groth SL, Ludwig C, Wang SY. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Netw Open*. 2023;6:e2330320.
90. Wang H, Masselos K, Tong J, Connor HRM, Scully J, Zhang S, Rafla D, Posarelli M, Tan JCK, Agar A, Kalloniatis M, Phu J. ChatGPT for addressing patient-centered frequently asked questions in glaucoma clinical practice. *Ophthalmol Glaucoma*. 2024;8:157.
91. Wu G, Zhao W, Wong A, Lee DA. Patients with floaters: answers from virtual assistants and large language models. *Digit Health*. 2024;10:20552076241229933.
92. Cappellani F, Card KR, Shields CL, Pulido JS, Haller JA. Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients. *Eye (Lond)*. 2024;38:1368-1373.
93. Lim ZW, Pushpanathan K, Yew SME, Lai Y, Sun CH, Lam JSH, Chen DZ, Goh JHL, Tan MCJ, Sheng B, Cheng CY, Koh VTC, Tham YC. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *EBioMedicine*. 2023;95:104770.
94. Reyhan AH, Mutaf Ç, Uzun İ, Yüksekayla F. A performance evaluation of large language models in keratoconus: a comparative study of ChatGPT-3.5, ChatGPT-4.0, Gemini, Copilot, Chatsonic, and Perplexity. *J Clin Med*. 2024;13:6512.
95. Raja H, Munawar A, Mylonas N, Delsoz M, Madadi Y, Elahi M, Hassan A, Abu Serhan H, Inam O, Hernandez L, Chen H, Tran S, Munir W, Abd-Alrazaq A, Yousefi S. Automated category and trend analysis of scientific articles on ophthalmology using large language models: development and usability study. *JMIR Form Res*. 2024;8:e52462.
96. Deiner MS, Deiner NA, Hristidis V, McLeod SD, Doan T, Lietman TM, Porco TC. Use of large language models to assess the likelihood of epidemics from the content of tweets: infodemiology study. *J Med Internet Res*. 2024;26:e49139.
97. Wang SY, Huang J, Hwang H, Hu W, Tao S, Hernandez-Boussard T. Leveraging weak supervision to perform named entity recognition in electronic health records progress notes to identify the ophthalmology exam. *Int J Med Inform*. 2022;167:104864.
98. Jaskari J, Sahlsten J, Summanen P, Moilanen J, Lehtola E, Aho M, Säpyskä E, Hietala K, Kaski K. DR-GPT: a large language model for medical report analysis of diabetic retinopathy patients. *PLoS One*. 2024;19:e0297706.
99. Shaheen A, Afflitto GG, Swaminathan SS. ChatGPT-assisted classification of postoperative bleeding following microinvasive glaucoma surgery using electronic health record data. *Ophthalmol Sci*. 2025;5:100602.
100. Aykut A, Sezenoz AS. Exploring the potential of code-free custom GPTs in ophthalmology: an early analysis of GPT store and user-creator guidance. *Ophthalmol Ther*. 2024;13:2697-2713.
101. Wu JH, Nishida T, Moghimi S, Weinreb RN. Effects of prompt engineering on large language model performance in response to questions on common ophthalmic conditions. *Taiwan J Ophthalmol*. 2024;14:454-457.
102. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol*. 2023;38:503-507.
103. Marshall R, Xu H, Dalvin LA, Mishra K, Edalat C, Kirupaharan N, Francis JH, Berkenstock M. Accuracy and completeness of large language models about antibody-drug conjugates and associated ocular adverse effects. *Cornea*. 2024;44:851-855.

104. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med*. 2023;29:1930-1940.
105. 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 (Lond)*. 2024;38:1252-1261.
106. Clusmann J, Kolbinger FR, Muti HS, Carrero ZI, Eckardt JN, Laleh NG, Löffler CML, Schwarzkopf SC, Unger M, Veldhuizen GP, Wagner SJ, Kather JN. The future landscape of large language models in medicine. *Commun Med (Lond)*. 2023;3:141.
107. U.S. Food and Drug Administration. 24 Hour Summary of the Digital Health Advisory Committee November 20-21, 2024, Available from: <https://www.fda.gov/media/184078/download>
108. Goutam B, Hashmi MF, Geem ZW, Bokde ND. A comprehensive review of deep learning strategies in retinal disease diagnosis using fundus images. *IEEE Access*. 2022;10:57796-57823.
109. Song S, Peng K, Wang E, Liu TYA. Enhancing large language model performance on ophthalmology board-style questions with retrieval-augmented generation. *Invest Ophthalmol Vis Sci*. 2025;66:3930.
110. Peng K, Wang E, Song S, Liu TYA. Leveraging retrieval-augmented generation with large language models in answering image-based board-style ophthalmology questions. *Invest Ophthalmol Vis Sci*. 2025;66:3931.
111. Wang E, Song S, Peng K, Liu TYA. Performance of large language models in answering questions regarding social determinants of health in ophthalmology. *Invest Ophthalmol Vis Sci*. 2025;66:3936.
112. Meyer A, Streichert T. Twenty-five years of progress-lessons learned from JMIR publications to address gender parity in digital health authorships: bibliometric analysis. *J Med Internet Res*. 2024;26:e58950.
113. Holman L, Stuart-Fox D, Hauser CE. The gender gap in science: How long until women are equally represented? *PLoS Biol*. 2018;16:e2004956.
114. Shah SS. Gender bias in artificial intelligence: empowering women through digital literacy. *Premier Journal of Artificial Intelligence*. 2024;1:1000088.
115. Brück O. A bibliometric analysis of geographic disparities in the authorship of leading medical journals. *Commun Med (Lond)*. 2023;3:178.
116. Chen H, Rider CI, Jurgens D, Teplitzkiy M. Geographical disparities in navigating rejection in science drive disparities in its file drawer. *Journal of Criminal Justice Education*. 2024;45.
117. Navigli R, Conia S, Ross B. Biases in large language models: origins, inventory, and discussion. *J Data Inf Qual*. 2023;15:1-21.
118. Rahkovsky I, Toney A, Boyack KW, Klavans R, Murdick DA. AI research funding portfolios and extreme growth. *Front Res Metr Anal*. 2021;6:630124.