

Systematic Review

Accuracy and Effectiveness of AI-Powered Systems in Patient Counseling, Education, and Management in Optometry and Related Eye-Care Settings: A Systematic Review

Manal M. Alharbi ^{1,*} , Emtenan M. Alharbi ^{2,*} , Zainab Ali Al-Hakmani ¹, Amr A. Arafat ^{3,4} , Melaf Alotaibi ⁵, Raneem Alatawi ⁶, Hanan Asiri ^{3,4} and Amani Aljurayyad ¹

¹ Department of Optometry, College of Applied Medical Sciences, King Saud University, Riyadh 11451, Saudi Arabia; zainabah@squ.edu.om (Z.A.A.-H.); amaniopto@gmail.com (A.A.)

² Clinical Pharmacy Department, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia

³ Research Department, Prince Sultan Cardiac Center, Riyadh 12611, Saudi Arabia; amr.arafat@med.tanta.edu.eg (A.A.A.); hananasiri.sa@gmail.com (H.A.)

⁴ Research and Innovation Institute, Ministry of Defense Health Services, Riyadh 12354, Saudi Arabia

⁵ College of Medicine, Imam Mohammad Ibn Saud Islamic University, Riyadh 11564, Saudi Arabia; melafalotaibi25@gmail.com

⁶ College of Medicine, King Saud University, Riyadh 11451, Saudi Arabia; falraneem@gmail.com

* Correspondence: amanal@ksu.edu.sa (M.M.A.); emtalharbi@ksu.edu.sa (E.M.A.)

Highlights

What are the main findings?

- This systematic review synthesized 39 eligible studies evaluating AI-supported patient counseling, education, communication, referral/follow-up, and management in optometry and related eye-care settings.
- Patient-facing LLMs and chatbot-based systems were prominent, particularly ChatGPT, Gemini/Bard, ChatGPT-4o, and disease-specific eye-care agents.

What are the implications of the main findings?

- Accuracy and readability were commonly evaluated, but patient-centered outcomes such as trust, satisfaction, usability, adherence, safety, and real-world clinical impact remained inconsistently reported.
- Current evidence supports AI as a supervised clinical adjunct rather than a replacement for optometrists, ophthalmologists, or clinician-led counseling.

Abstract

Background/Objectives: Artificial intelligence (AI), including large language models, chatbots, machine-learning systems, and hybrid tools, is increasingly used to support patient-facing eye-care communication. This systematic review evaluated the accuracy and effectiveness of AI-powered systems used for patient counseling, education, communication, referral/follow-up, and management support in optometry and related eye-care settings. **Methods:** A systematic review was conducted according to a predefined protocol and PRISMA 2020 reporting principles. Searches covered studies published from January 2020 to 31 May 2026 in PubMed, Embase, Web of Science, Cochrane Library, and IEEE Xplore. Eligible studies were original primary studies evaluating AI-supported tools for patient-facing counseling, education, question answering, treatment or medication guidance, triage, referral, follow-up, screening linked to management, or clinical decision support. Methodological quality was appraised using relevant JBI critical appraisal tools. **Results:** Thirty-nine studies were included in the qualitative synthesis. JBI appraisal indicated heterogeneous study designs and reporting quality, with common limitations related



Academic Editors:

Andrzej Grzybowski, Kai Jin and
Polat Göktas

Received: 21 June 2026

Revised: 18 July 2026

Accepted: 22 July 2026

Published: 24 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

to simulated prompts or AI-output evaluations, variable comparators, and inconsistent outcome reporting. Evidence was dominated by patient-facing large language models and chatbot-based tools used for patient education, question answering, readability improvement, glaucoma and myopia counseling, diabetic retinopathy referral/follow-up, cataract education, oculoplastic and retinal-condition questions, and multilingual educational support. Study designs, AI models, prompting approaches, comparators, clinical topics, and outcome definitions varied widely, supporting narrative synthesis rather than quantitative pooling. **Conclusions:** AI-powered systems show potential as supervised adjunctive tools for eye-care counseling, education, and management-related communication. However, evidence remains heterogeneous and dependent on simulated prompts, AI-generated outputs, and model-based evaluations. Future research should prioritize standardized evaluation, real-patient validation, safety monitoring, readability control, and patient-centered outcomes before routine implementation in optometry and broader eye-care practice across diverse clinical settings.

Keywords: artificial intelligence; optometry; patient counseling; patient education; ophthalmology; large language models; ChatGPT; systematic review

1. Introduction

Artificial intelligence is increasingly integrated into healthcare, with applications ranging from diagnostic support to patient education, communication, and long-term disease management [1–5]. In optometry and related eye-care practice, patient counseling is essential because many eye conditions require ongoing monitoring, treatment adherence, lifestyle education, symptom recognition, and timely referral when warning signs appear [3,4]. Recent eye-care studies in this area have examined AI-supported referral/follow-up and management-support workflows [6–18], other patient-facing communication applications [19–37], and recent patient education and question-answering studies [38–44]. AI-powered tools may assist by generating explanations, answering common questions, producing patient handouts, simplifying medical terminology, and reinforcing information delivered during clinical encounters [38–44].

Early eye-care AI research demonstrated strong performance for image-based disease screening, particularly diabetic retinopathy detection [1,2]. Although this diagnostic literature established the technical feasibility of AI in eye care, it does not directly answer whether AI-generated patient-facing information is accurate, readable, safe, and clinically appropriate for counseling [1–4]. Patient-facing AI requires a different standard of evaluation because outputs must be understandable to patients and must not delay or replace professional care [38–50].

Recent large-language-model studies have expanded from diagnostic or examination-style tasks toward patient education, online question answering, glaucoma counseling, pediatric ophthalmology education, multilingual materials, and real patient questions [38–50]. These studies suggest that AI tools may generate useful information, but their performance varies by model, prompt, language, clinical topic, comparator, and scoring method [38–44]. Safety, readability, and patient-centered outcomes remain major concerns [38–50].

Although previous systematic and scoping reviews have examined broad uses of large language models and artificial intelligence in ophthalmology, most have focused on general ophthalmic applications, diagnostic or educational use cases, or broad LLM performance rather than the specific patient-facing pathway addressed here. The present review therefore provides a distinct synthesis of original primary studies evaluating AI-supported

counseling, education, communication, referral/follow-up, and management-related support in optometry and related eye-care settings.

This systematic review focuses on AI-powered systems used for patient counseling, education, communication, referral/follow-up, and management support in optometry and related eye-care settings [38–50]. Because optometric practice overlaps with broader ophthalmic patient education, ophthalmology studies were included when the AI application was patient-facing and addressed counseling, education, question answering, referral/follow-up, or management support rather than diagnostic image classification alone [45–50].

The review question was: What is the accuracy and effectiveness of AI-powered systems used to support patient-facing counseling, education, communication, referral/follow-up, and management-related tasks in optometry and related eye-care settings? Accordingly, this systematic review aimed to synthesize original primary evidence on these patient-facing AI applications and to evaluate their reported accuracy, clinical quality, readability, safety, and patient-centered or implementation outcomes, when available.

Clinical Significance

The clinical significance of this topic lies in the potential use of AI as a supplementary counseling and education tool [38–44]. If accurate, readable, and appropriately supervised, AI-generated responses may help reinforce information provided during clinical visits, answer common patient questions, support follow-up instructions, and improve access to patient education materials [38–44]. At the same time, patient-facing AI may also introduce risk if outputs are incomplete, overly reassuring, difficult to understand, or not aligned with referral urgency [45–50]. For this reason, a review focused specifically on patient-facing communication, rather than diagnostic performance alone, is clinically relevant for optometrists, ophthalmologists, and teams considering AI-assisted education or follow-up support.

2. Materials and Methods

2.1. Study Design and Reporting Framework

This study was conducted as a systematic review using a predefined protocol. Reporting was guided by PRISMA 2020 principles [51], and methodological appraisal was guided by relevant Joanna Briggs Institute (JBI) critical appraisal tools and evidence-synthesis guidance [52]. The review protocol was prospectively registered in PROSPERO (registration number: CRD420251234880). Eligibility criteria, information sources, screening procedures, extraction fields, outcomes, and synthesis methods were defined before the review was conducted. The PRISMA 2020 checklist and complete search strategy are provided in Supplementary File S1.

Protocol amendments were recorded in PROSPERO. These updates mainly refined the title, citation and author details, rationale, domain terms, review objectives, eligibility framework, outcome wording, quality-appraisal approach, search date range, and synthesis plan. The main methodological updates included extending the search date to 31 May 2026, aligning the methodological appraisal approach with JBI tools, and using narrative synthesis rather than meta-analysis because the included studies were highly heterogeneous in AI models, prompts, comparators, clinical tasks, rating scales, and outcome definitions. These updates were made to align the registered record with the final patient-facing scope and eligible evidence base, and they did not change the core focus of the review on patient-facing AI applications in optometry and related eye-care settings.

2.2. Information Sources and Search Strategy

The literature search covered studies published from January 2020 to 31 May 2026. Information sources included PubMed, Embase, Web of Science, Cochrane Library, and IEEE Xplore. The search strategy was developed from the eligibility framework using three structured concept blocks: (i) eye-care population or context terms, including optometry, ophthalmology, ocular disease, and related eye-care settings; (ii) AI-intervention terms, including artificial intelligence, machine learning, deep learning, large language models, and chatbots; and (iii) patient-facing function or outcome terms, including counseling, education, communication, question answering, referral, follow-up, management, readability, and understandability. Search terms were adapted to each database interface using appropriate keywords, controlled vocabulary, and Boolean combinations where applicable. Search reporting was aligned with PRISMA-S principles for documenting database-specific search strategies [53]. The complete search strategy is provided in Supplementary File S1. The May 2026 update was treated as an updated search to capture newly eligible studies, and the final 39-study synthesis reflects all records retained after full-text eligibility reassessment and search updating.

2.3. Eligibility Criteria

Eligible studies included original primary studies evaluating AI-supported tools in optometry or ophthalmology for patient counseling, patient education, healthcare communication, frequently asked questions, treatment or medication guidance, triage, referral, follow-up, screening linked to patient management, or patient-facing clinical decision support. Eligible AI systems included LLMs, chatbots, machine-learning tools, deep-learning systems, rule-based systems, and hybrid systems. Randomized trials, observational studies, cross-sectional studies, pilot studies, quasi-experimental studies, case studies, case series, and comparative evaluation studies were considered when they reported relevant patient-facing or management-related outcomes.

Reviews, editorials, letters, response articles, survey/review-style papers without primary outcome data, non-eye-care AI studies, examination/training-only studies, and purely technical image classification, segmentation, or prediction studies without patient-facing counseling, education, referral, follow-up, or management outcomes were excluded.

No restriction was applied to the language of AI outputs or patient-facing materials. Studies evaluating non-English or multilingual patient-facing materials were eligible when sufficient information was available in English, or could be reliably interpreted, to assess eligibility and extract the relevant methods, outcomes, and results. Studies were not included if language barriers prevented reliable assessment of eligibility or extraction of outcome data.

Referral, triage, screening, or management tools were included only when the AI application directly supported patient-facing communication, referral, follow-up, or management decisions. Tools used only for clinician-facing technical prediction, segmentation, or classification without patient-facing pathway relevance were excluded.

Studies based on real patient questions or interactions, simulated patient prompts or vignettes, patient education materials, and expert-only AI-output evaluations were eligible when they met the patient-facing or management-related criteria. The eligibility framework is summarized in Table 1.

Table 1. Eligibility framework.

Domain	Eligibility Definition
Population	Ophthalmic or optometric patients, caregivers, simulated patient profiles, patient questions, patient-facing educational materials, or eye-care clinical contexts.
Intervention	AI-powered systems, including LLMs, chatbots, machine-learning tools, deep-learning models, rule-based systems, or hybrid systems used for counseling, education, communication, referral, follow-up, or management support.
Comparator	Human counseling, clinician or expert assessment, professional guidelines, society patient education materials, reference standards, or other AI systems.
Outcomes	Accuracy, readability, completeness, appropriateness, reliability, usefulness, satisfaction, trust, usability, safety, potential harm, actionability, and other patient-centered or communication-related outcomes.
Exclusions	Reviews, commentaries, letters, response articles, non-eye-care AI studies, examination-only studies, and purely technical classification/segmentation/prediction studies without patient-facing relevance.

2.4. Study Selection, Data Extraction, and Methodological Quality Appraisal

Records were managed using Covidence systematic review software (Veritas Health Innovation, Melbourne, Australia; available at www.covidence.org). Titles, abstracts, and full texts were assessed against the eligibility criteria by two reviewers independently, with disagreements resolved through discussion or consultation with a third reviewer when consensus was not reached. Data extraction and JBI methodological appraisal were also completed using the same two-reviewer and consensus approach. A formal inter-rater agreement statistic, such as Cohen's kappa, was not calculated because all discrepancies were resolved by consensus and the review did not use a single quantitative eligibility score.

Data were extracted using a standardized, predefined extraction form implemented as a structured, wide-format spreadsheet with an extraction dictionary. Extracted variables included study identification, publication year, country, study design, setting, participant or evaluation type, AI model name and type, comparator, patient-facing function, reported outcomes, exact outcome values where available, and source traceability. When a study evaluated multiple AI systems, each model was captured separately for model-family summaries, while the study itself remained counted once in the final synthesis.

Methodological quality was appraised using the relevant JBI critical appraisal tool for each study design. Appraisal outcomes were not used as exclusion criteria; instead, they informed the narrative interpretation of confidence in the evidence. Studies with randomized, real-patient, implementation, or clearly reported comparative designs were interpreted as providing stronger support for applied conclusions, whereas studies based on brief reports, simulated prompts, AI-output-only evaluations, unclear sampling, or limited comparator reporting were interpreted more cautiously. Common appraisal concerns included heterogeneous study designs, limited real-patient validation, reliance on simulated prompts or AI-output evaluations, and inconsistent reporting of comparators, prompts, model versions, and patient-centered outcomes. Accordingly, output-level findings on accuracy, readability, or appropriateness were not interpreted as direct evidence of patient understanding, adherence, safety, or real-world clinical effectiveness unless these outcomes were specifically measured.

2.5. Outcomes

The primary outcome was the accuracy or clinical quality of AI-generated counseling or educational information, including correctness, completeness, appropriateness, safety, or alignment with clinician/expert judgment, professional guidelines, society educational materials, or other reference standards. Secondary outcomes included readability, reliability,

efficiency, comprehensiveness, understandability, usefulness, satisfaction, trust, usability, actionability, safety, potential harm, adherence, and other patient-centered outcomes when reported.

2.6. Statistical Analysis and Synthesis

A descriptive synthesis was performed using the final eligible study set ($n = 39$). Studies were grouped by primary AI model or system category, primary patient-facing function, clinical topic, comparator type, reported outcome domain, and JBI appraisal concern level. Categorical variables were summarized descriptively, with 39 used as the denominator when study-level proportions were reported. Outcome-reporting frequencies and model-family summaries were based on the final 39-study dataset. Performance outcomes were extracted exactly as reported by each study.

Because the included studies differed substantially in AI model type, prompt structure, comparator, clinical topic, rating scale, and outcome measurement, meta-analysis or quantitative pooling was not appropriate. Findings were therefore synthesized narratively using tables and figures. Formal certainty-of-evidence rating using the GRADE approach was not undertaken because the heterogeneity in study designs, AI models, comparators, outcome definitions, and reference standards, together with the predominance of simulated-prompt and AI-output evaluations rather than comparative patient outcomes, precluded the outcome-specific evidence syntheses required for GRADE. Certainty was therefore considered narratively alongside the JBI appraisal.

The study-selection process followed the PRISMA stages of identification, screening, retrieval, eligibility assessment, and inclusion. Eligibility criteria and screening procedures are described in Sections 2.3 and 2.4, whereas the final study-selection counts are reported in Section 3 and summarized in Figure 1.

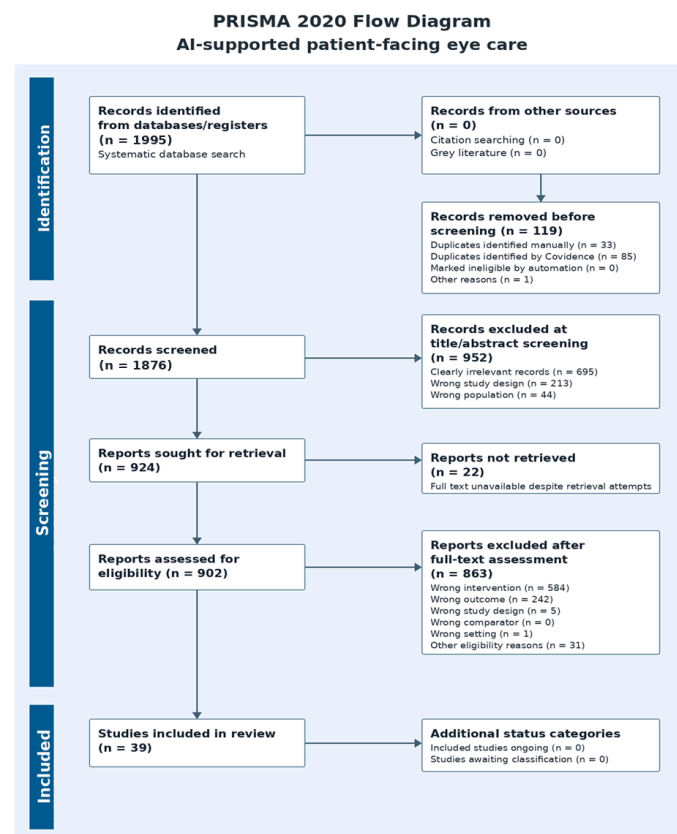


Figure 1. PRISMA flow diagram of study selection for the systematic review of AI-supported patient-facing education, counseling, communication, and management in optometry and ophthalmology.

3. Results

3.1. Study Selection

A total of 1995 records were identified from bibliographic databases. Before screening, 119 records were removed, including 33 duplicates identified manually, 85 duplicates identified by Covidence, and one record removed for other reasons. No trial registries were searched; all identified records came from bibliographic database searches.

After removal of these records, 1876 records were screened. At title and abstract screening, 952 records were excluded: clearly irrelevant records ($n = 695$), wrong study design ($n = 213$), and wrong population ($n = 44$). A total of 924 reports were sought for retrieval, and 22 reports were not retrieved because the full text was unavailable despite retrieval attempts. Of 902 reports assessed for eligibility, 863 were excluded after full-text assessment because of wrong intervention ($n = 584$), wrong outcome ($n = 242$), wrong study design ($n = 5$), wrong comparator ($n = 0$), wrong setting ($n = 1$), or other eligibility reasons ($n = 31$).

Thirty-nine studies were included in the final qualitative synthesis. The study-selection process is summarized in Figure 1 using the PRISMA stages of identification, screening, retrieval, eligibility assessment, and inclusion. The PRISMA flow diagram was generated using Covidence, and the narrative text reports the final selection counts used for the 39-study synthesis. The 39 included studies correspond to the studies cited in the main reference list and listed in Supplementary Table S1.

3.2. Characteristics of the Final Included Studies

The final synthesis included 39 studies and centered on patient-facing AI applications in optometry and ophthalmology, including patient education, question answering, counseling support, patient handout generation, readability improvement, referral/follow-up support, and management-related communication. Many included studies evaluated LLMs or chatbot-based systems, while others examined AI-enabled triage, referral, or management-support tools linked to patient care pathways. This final set reflects the updated search and final eligibility reassessment rather than a preliminary Covidence export.

To avoid duplication with the reference list, detailed study identifiers and DOI/stable identifier information are retained in Supplementary Table S1, while the main text summarizes the evidence domains, AI model categories, patient-facing functions, outcome domains, and methodological appraisal findings.

The evidence base expanded mainly after 2023, reflecting rapid growth in LLM-based eye-care communication and patient-education research. The studies included a mixture of online LLM-output evaluations, simulated patient questions, patient education materials, real patient questions, randomized or comparative evaluations, and referral/follow-up or management-support studies. Because designs and reporting standards varied, results were synthesized descriptively rather than pooled statistically. The publication-year distribution of the included studies is shown in Figure 2.

The included studies were distributed as follows: one study in 2022, three studies in 2023, 14 studies in 2024, 17 studies in 2025, and four studies in 2026. The 2026 count reflects studies identified up to 31 May 2026 and does not represent a full calendar year.

To strengthen the narrative synthesis, the included studies were grouped by patient-facing application and AI model type. The main application groups included patient education and readability, question answering and counseling, referral/follow-up and management support, and multilingual or accessibility-focused communication. The main model groups included general-purpose LLMs or chatbots, AI screening/referral/follow-up workflows, and disease-specific or multimodal eye-care agents. These domains and examples from the included studies are summarized in Table 2.

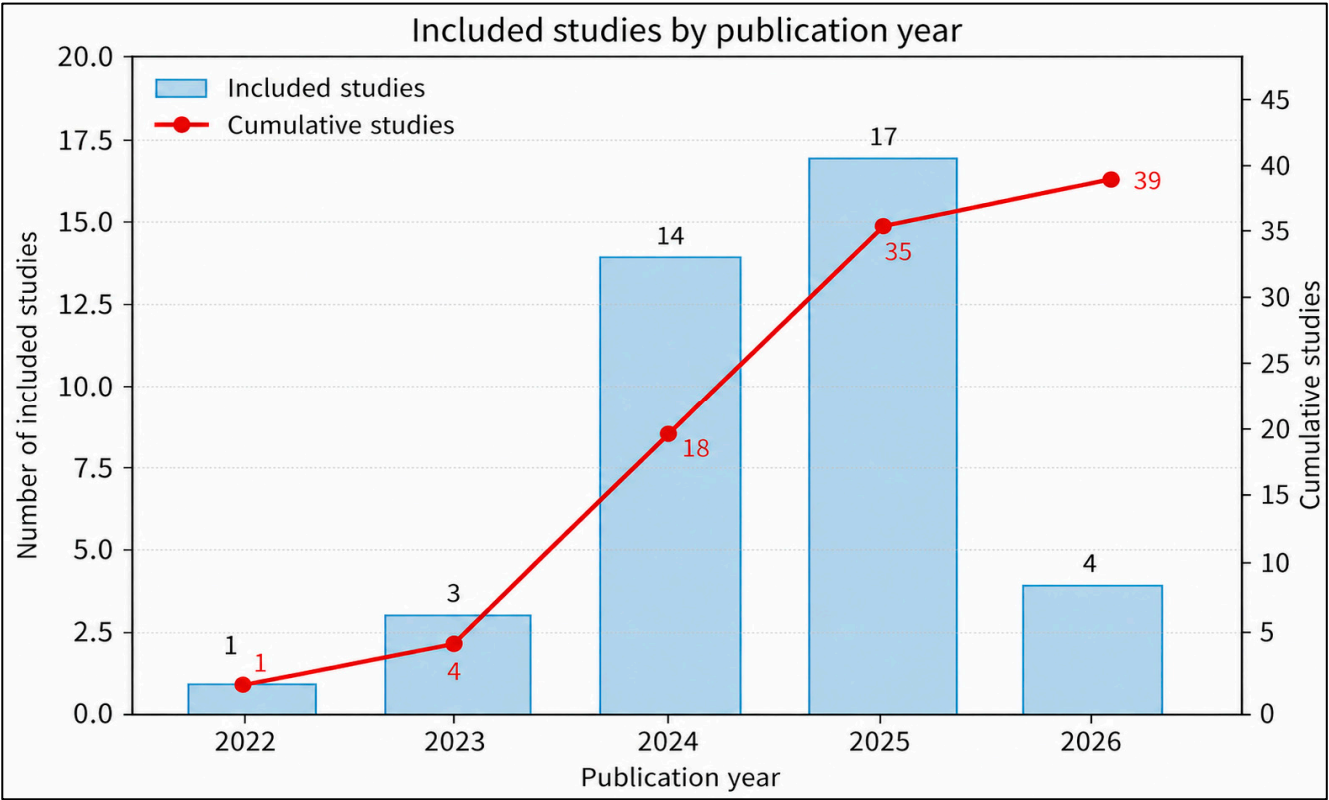


Figure 2. Distribution of included studies by publication year and cumulative study count in the final 39-study synthesis.

Table 2. Main evidence domains represented in the 39 included studies.

Domain	Examples in the Retained Evidence Base
Patient education and readability	Dry eye disease, cataract surgery, myopia, glaucoma, pediatric ophthalmology, AAO/AAPOS materials, medical-jargon translation, patient handouts.
Question answering and counseling	Cataract counseling, retinal detachment information, oculoplastic patient queries, choroidal melanoma support, real ophthalmology patient questions, glaucoma counseling, herpes zoster ophthalmicus education.
Referral, follow-up, and management support	At-risk referral from primary care to eye specialists, diabetic retinopathy screening/follow-up, AI-human hybrid teleophthalmology workflows, strabismus screening/management, geographic atrophy prediction/management.
Multilingual and accessibility-focused studies	Traditional Chinese glaucoma counseling, Turkish glaucoma education, Spanish-language patient handouts, simplification of clinical notes and medical terminology.
Model families	ChatGPT family models, Gemini/Bard, ChatGPT-4o, Microsoft Copilot/Bing, EyeGPT, ChatMyopia, FFA-GPT, and other ML or hybrid systems used in patient-facing pathways.

3.3. AI Models and Patient-Facing Functions

Large language models and chatbot-based systems were the most prominent primary AI category in the included evidence base, representing 24 of 39 studies (61.5%). AI screening, referral, follow-up, or management-support workflows accounted for

10 studies (25.6%), while disease-specific or multimodal eye-care agents accounted for five studies (12.8%). Frequently represented systems included ChatGPT-family models, Gemini/Bard, ChatGPT-4o, Microsoft Copilot/Bing, disease-specific AI agents, and ophthalmology-specific LLM or multimodal platforms. For Figure 3, each study was assigned to one primary AI model or system category based on the main AI application evaluated. The primary AI model or system categories are summarized in Figure 3.

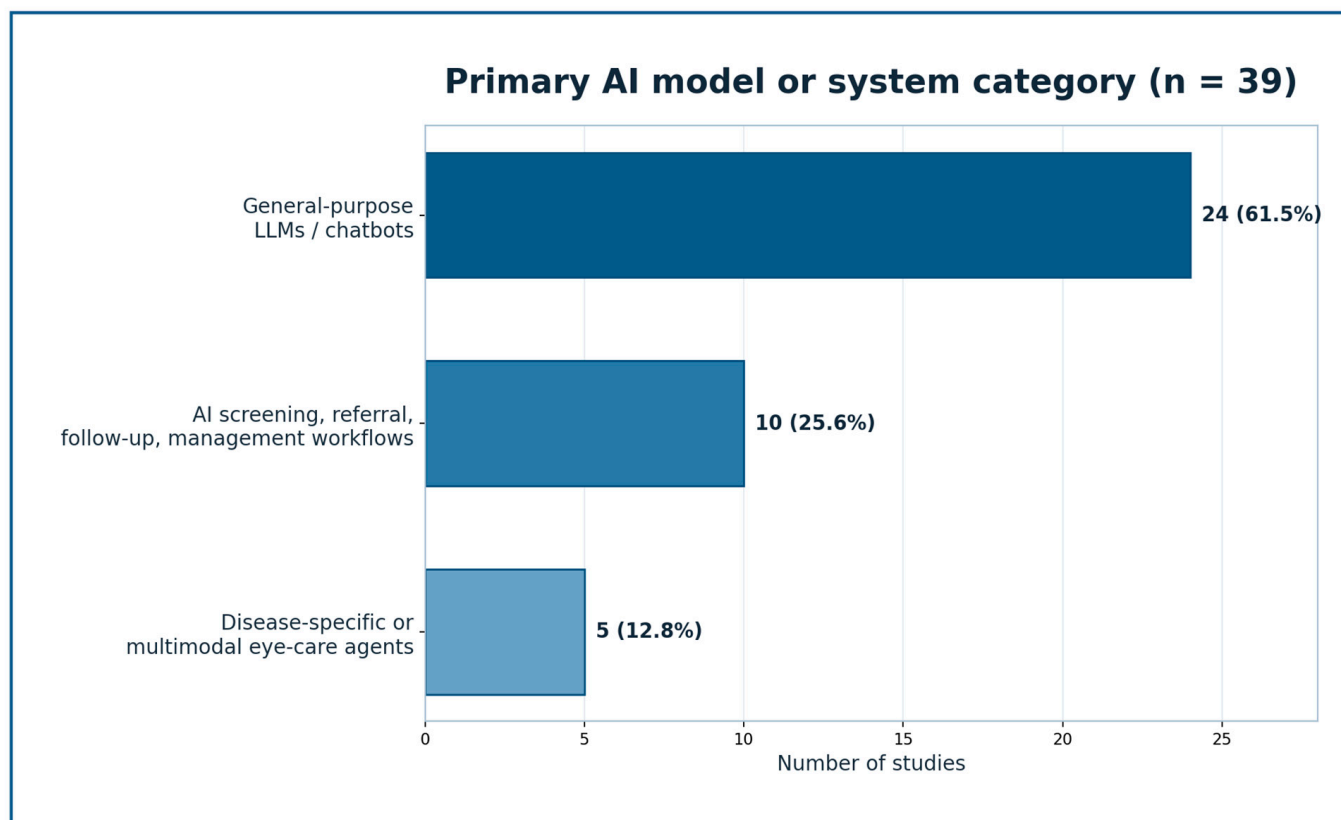


Figure 3. Primary AI model or system category represented in the final 39-study synthesis.

Patient-facing functions included answering common eye-care questions, generating patient educational materials, simplifying ophthalmic language, preparing plain-language summaries, supporting patient handouts, providing counseling-related responses, and supporting referral or follow-up workflows. The most common primary function was patient education, readability, or handout generation, reported in 16 of 39 studies (41.0%), followed by question answering and counseling in 12 studies (30.8%) and referral, follow-up, or management support in 11 studies (28.2%). For classification purposes, each study was assigned to one primary patient-facing function based on its main reported AI application; therefore, the counts in Figure 4 represent mutually exclusive study counts and sum to 39 studies. The retained management and screening studies were included only when their AI application was linked to referral, follow-up, or patient-facing management, not when the paper was limited to technical image classification alone. The primary patient-facing functions are summarized in Figure 4.

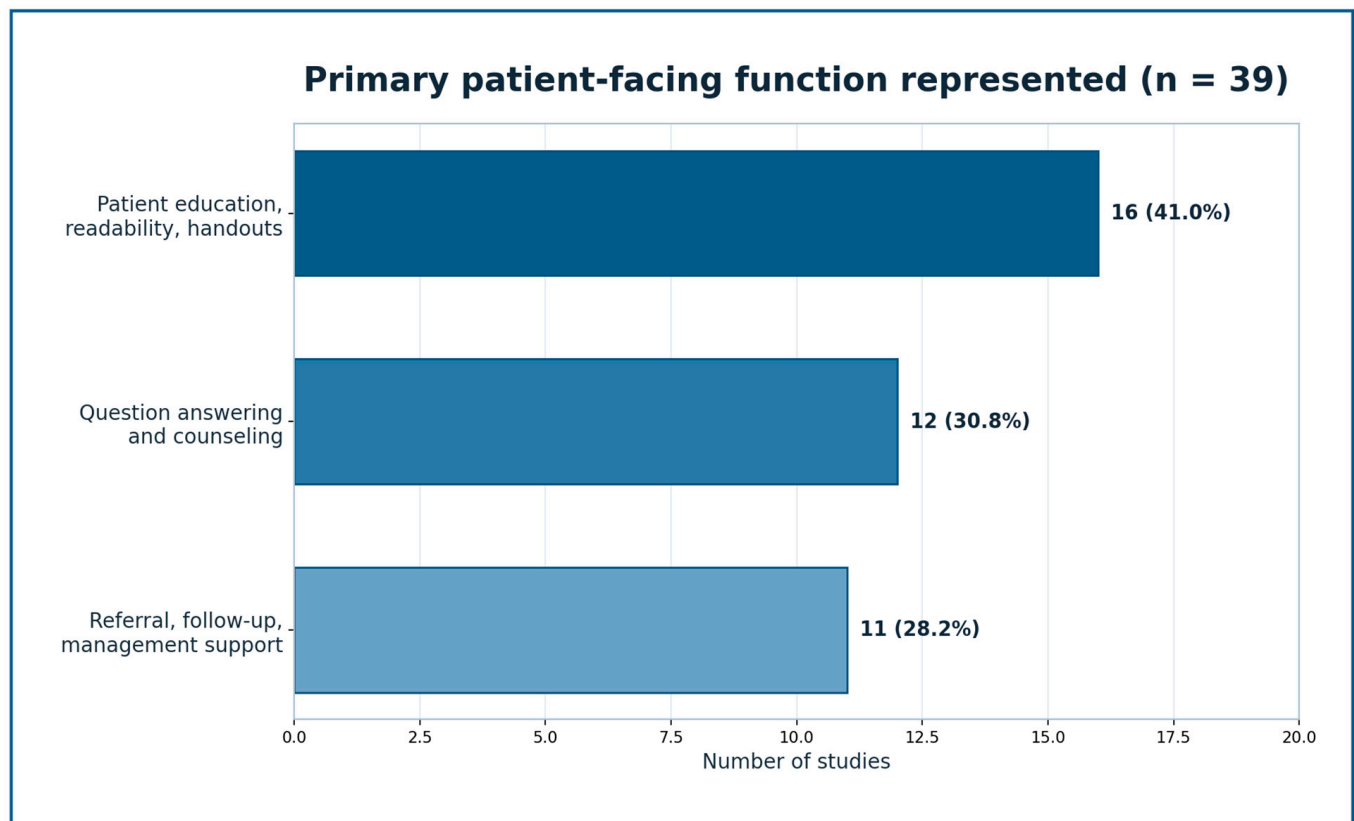


Figure 4. Primary patient-facing function represented in the final 39-study synthesis.

3.4. Outcome Reporting

Across the included 39 studies, the most emphasized outcome domains were accuracy, clinical quality, appropriateness, readability, completeness, understandability, and reliability of AI-generated information. Some studies also evaluated efficiency, actionability, usefulness, safety, potential harm, satisfaction, or trust. Because studies used different rubrics and did not apply a shared threshold for “acceptable” accuracy or readability, pooled acceptability estimates were not calculated. Instead, outcomes were interpreted as directional and domain-specific findings. Among the seven recent patient-facing LLM studies, six reported generally favorable or clinically acceptable accuracy, quality, or appropriateness with important qualifications, while all seven identified limitations related to readability, completeness, model dependence, prompt dependence, or the need for clinician review. Across the full 39-study evidence base, patient-centered outcomes such as satisfaction, trust, usability, adherence, behavioral change, and long-term clinical impact remained inconsistently reported. These outcomes were grouped narratively into clinical quality, patient communication, implementation, and safety domains, as summarized in Figure 5.

Figure 5 is intended as a narrative synthesis map rather than a pooled outcome model. Studies evaluating AI answers, patient handouts, or educational materials contributed mainly to the clinical-quality and patient-communication domains; studies evaluating referral, follow-up, triage, or screening-linked management contributed mainly to the implementation domain; and studies reporting misinformation, missing referral warnings, potentially harmful advice, or need for clinician oversight informed the safety domain. Because many publications assessed more than one domain, the domains should be interpreted as overlapping themes rather than mutually exclusive categories.

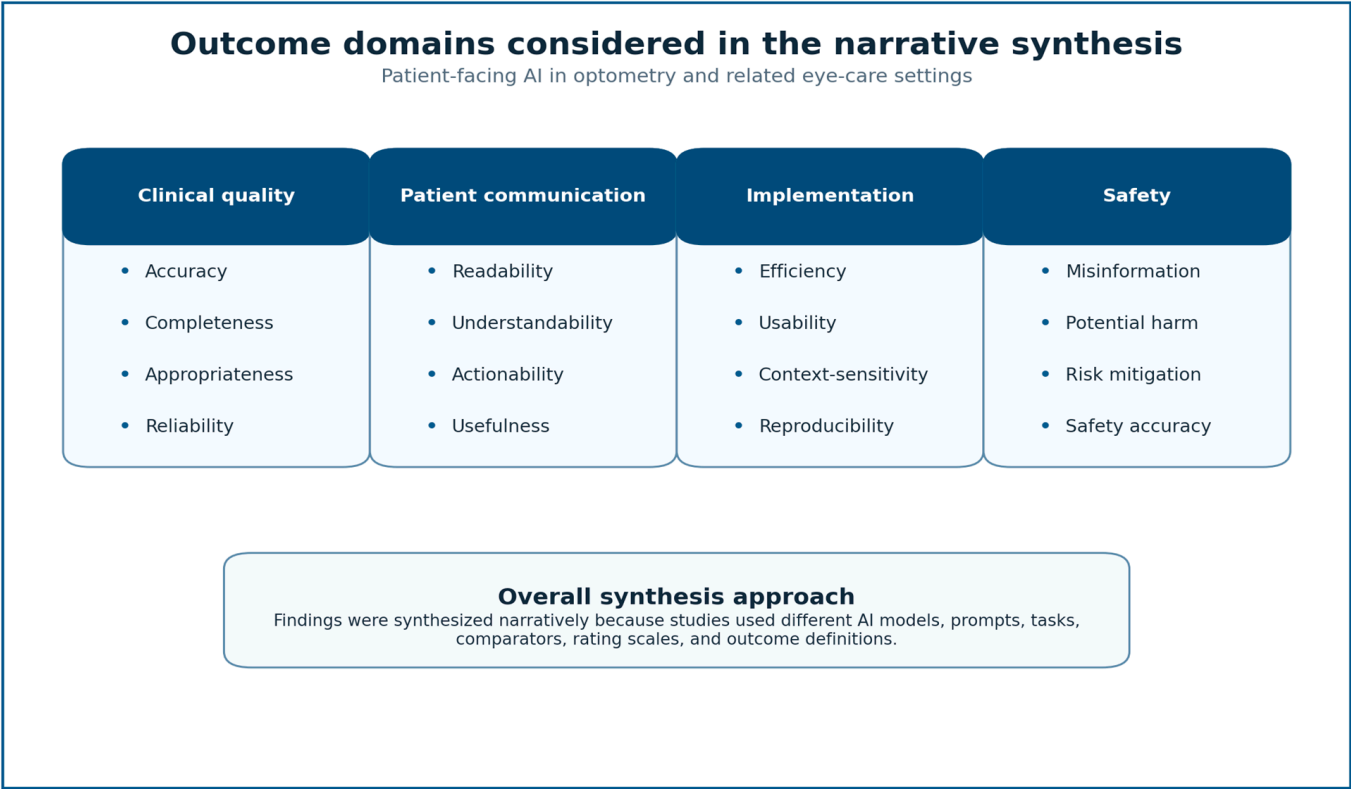


Table 3. Cont.

Evidence or Appraisal Dimension	Summary in the 39-Study Synthesis	Influence on Interpretation
Primary patient-facing function	Patient education, readability, or handout generation: 16/39 (41.0%); question answering and counseling: 12/39 (30.8%); referral, follow-up, or management support: 11/39 (28.2%).	Evidence was synthesized by function because counseling, education, and follow-up tools had different comparators and outcomes.
Outcome interpretation	Accuracy, clinical quality, appropriateness, readability, completeness, and reliability were common output-level outcomes; patient-centered and behavioral outcomes were reported less consistently.	Output-only or expert-evaluated findings were not treated as proof of real-world clinical effectiveness.
Implementation inference	Prospective patient deployment and long-term clinical outcomes were limited across the literature.	Recommendations for governance, safety monitoring, and real-patient validation were framed as future implementation requirements rather than proven effects.

Because the included studies differed substantially in design, task type, AI model, scoring rubric, and outcome scale, outcome patterns were synthesized narratively rather than pooled statistically. This approach avoided combining clinically heterogeneous accuracy, readability, usefulness, and safety outcomes into a single summary estimate.

3.5. Summary of Recent Patient-Facing LLM Studies

Several recent studies directly evaluated patient-facing LLM applications in ophthalmology and related eye-care settings. These studies are important because they move beyond general diagnostic AI and assess real patient questions, patient educational materials, readability, counseling quality, language adaptation, and clinical appropriateness. Table 4 provides a focused summary of seven recent patient-facing LLM studies published in 2025–2026, selected because they most directly represent contemporary counseling and education applications relevant to the review objective.

Table 4. Selected recent patient-facing AI studies included in the review.

Study	AI/Focus	Main Finding	DOI
Riazi Esfahani et al., 2025 [38]	Common ophthalmology questions	ChatGPT responses were generally favorable for accuracy/comprehensiveness/clarity, but readability remained above the recommended level for patient education.	10.7759/cureus.87920
Dal et al., 2026 [39]	Turkish glaucoma patient education	LLMs produced useful educational content, but readability and comprehensiveness differed by model and prompting approach.	10.14744/eer.2025.93723
Yang et al., 2025 [40]	Pediatric ophthalmologic surgery materials	AI-generated pediatric surgical education materials were not consistently superior to AAPOS materials on quality domains.	10.3928/01913913-20250404-01

Table 4. *Cont.*

Study	AI/Focus	Main Finding	DOI
Bondok et al., 2025 [41]	AI chatbot versus ophthalmologist responses	ChatGPT-4o responses were comparable in accuracy and empathy, but tended to be longer and less readable than clinician responses.	10.2147/OPTH.S549820
Lin et al., 2026 [42]	Glaucoma counseling in Traditional Chinese	ChatGPT and Gemini gave reasonably appropriate responses, with ChatGPT scoring higher overall than Gemini.	10.4103/tjo.TJO-D-25-00058
Wu et al., 2026 [43]	ChatGPT-4o patient handouts versus AAO materials	ChatGPT-4o generated handouts competitive with AAO materials, including notable findings in Spanish-language materials.	10.1167/tvst.15.2.14
Hamzeh et al., 2026 [44]	Real ophthalmology patient questions	Most ChatGPT-4o responses were accurate and complete, but some were incomplete or unacceptable, and readability remained high.	10.1016/j.xops.2025.101007

3.6. Narrative Synthesis of Reported Effectiveness and Limitations

Overall, the included studies suggest that AI-powered systems can generate useful, coherent, and often clinically relevant patient-facing information in eye-care contexts. However, performance was not uniform. The quality of AI-generated output varied by clinical topic, model version, prompt structure, language, reference standard, and evaluation criteria. Some studies reported favorable accuracy, appropriateness, or usefulness, whereas others identified incomplete, overly complex, potentially unsafe, or less readable outputs.

Across the included studies, measurable patient-level adverse events or observed unsafe behaviors were rarely reported. Reported drawbacks were therefore categorized narratively as potential misinformation, incomplete or overly complex responses, missing safety-netting or referral warnings, readability above patient education levels, and the need for clinician review before patient use.

Recent LLM studies support the potential role of AI as a drafting, simplification, or educational support tool. AI-generated patient handouts and answers to common patient questions may be useful, particularly when reviewed by clinicians. However, readability remained a recurring concern because some AI-generated responses were written above recommended patient education levels. In multilingual contexts, AI systems showed promise, but performance was not consistently superior to professional or society-developed patient education materials.

Chronic eye disease education, glaucoma counseling, diabetic retinopathy referral/follow-up, myopia education, cataract surgery counseling, and oculoplastic or retinal-condition patient questions were recurrent areas of evaluation. These topics are clinically important because patient understanding, adherence, symptom recognition, and follow-up behavior influence outcomes. The evidence supports AI as a supervised support tool, not as an autonomous source of clinical advice.

4. Discussion

This systematic review synthesized 39 eligible studies evaluating AI-powered systems for patient counseling, education, communication, referral/follow-up, and management-related support in optometry and related eye-care settings. The main message is not simply that AI can answer ophthalmic questions; rather, the evidence shows that AI performance depends strongly on the clinical task, model version, prompt, language, comparator, and

outcome measure. Across the included studies, large language models and chatbot-based tools showed potential for producing patient-facing explanations and educational materials, but the evidence remains too heterogeneous to support unsupervised clinical use.

A focused interpretation is important because the objective of this review is different from broad AI-in-ophthalmology reviews. Diagnostic screening studies and technical deep-learning papers have shown that AI can perform well in image-based eye-care tasks, especially diabetic retinopathy detection [1,2]. However, patient counseling requires different evidence. A system used for counseling must provide correct information, communicate at an appropriate reading level, avoid unsafe reassurance, recognize when referral is needed, and support patient understanding without replacing professional judgment. Therefore, this review emphasizes patient-facing communication outcomes rather than technical diagnostic performance alone.

4.1. Principal Interpretation of the 39-Study Evidence Base

The retained evidence base suggests that AI may be useful as a supervised adjunct for common eye-care communication tasks. These tasks include answering frequently asked patient questions, preparing or simplifying educational materials, explaining chronic diseases such as glaucoma, diabetic retinopathy, myopia, and dry eye disease, supporting cataract and pediatric ophthalmology education, and reinforcing referral or follow-up instructions. This is clinically relevant because many eye-care conditions require repeated counseling over time, and patients often seek explanations outside the clinic through online resources.

The strength of the evidence is strongest for model-output evaluation and educational content generation, and weakest for real-world patient outcomes. Many studies assessed whether AI responses were accurate, complete, readable, or appropriate when judged by clinicians, experts, guidelines, or society patient information. Fewer studies examined whether patients actually understood the output, trusted it appropriately, followed advice safely, adhered to treatment, or attended follow-up. For this reason, the current review supports AI as an educational support tool, not as an independent counseling provider.

The findings also show that accuracy alone is not enough. An AI response may be factually acceptable but too long, too complex, insufficiently personalized, or missing a safety warning. Conversely, a readable response may omit important clinical nuance. Patient-facing quality therefore requires a combined assessment of accuracy, completeness, readability, actionability, safety, and clinical appropriateness. This combined standard is especially important in optometry and ophthalmology because symptoms such as flashes, floaters, painful red eye, sudden visual loss, postoperative symptoms, and worsening diabetic eye disease may require urgent referral rather than generic reassurance.

4.2. Comparison with Previous Literature and Reviews

The present review is consistent with the earlier ophthalmology AI literature in showing that AI has moved rapidly from diagnostic tasks toward broader clinical communication and education. Keskinbora and Guven and Costin et al. described the expanding role of AI in ophthalmology while emphasizing limitations related to transparency, reliability, ethics, and clinical implementation [3,4]. The present review extends that background by focusing on the patient-facing layer of care: whether AI-generated information can support counseling, education, question answering, follow-up communication, or management-related advice.

Several recent reviews of LLMs in ophthalmology and eye education also support the need for a cautious interpretation. Zhang et al. mapped LLM evaluations in ophthalmology and described substantial variability in models, tasks, subspecialties, and outcome measures [45]. Agnihotri et al. similarly highlighted rapid growth in

ophthalmology LLM publications but noted that evaluation methods remain inconsistent [46]. Artiaga et al. described potential utility for clinicians, patients, researchers, and educators, while Li et al., Andrew and Rodenko, and See et al. emphasized the need for validation, safety assessment, and clearer reporting before routine use in eye-care education [47–50]. The current review adds a narrower contribution by separating patient-facing counseling and education from broad diagnostic or technical AI applications.

The patient-facing primary studies included in this review are also broadly consistent with the review literature. Studies of common ophthalmology questions, glaucoma education, pediatric ophthalmology materials, AI-generated handouts, and ophthalmologist-versus-chatbot responses suggest that AI outputs can be useful, but not uniformly reliable or patient-ready [38–44]. Some studies found favorable accuracy or completeness, whereas others reported concerns about readability, length, quality variation, missing warnings, or inferior performance compared with society-developed educational materials. This pattern supports a balanced conclusion: AI can assist clinicians in preparing or reviewing educational content, but outputs should be checked before they are given to patients. Table 5 summarizes how the present review compares with the key previous literature and selected patient-facing studies.

Table 5. Comparison of the present review with the key previous literature and patient-facing studies.

Reference	Main Contribution	Comparison with This Review
Gulshan et al. [1]	AI performed strongly for diabetic retinopathy image screening.	Useful background only; counseling safety needs separate evidence.
Abramoff et al. [2]	Autonomous diabetic retinopathy detection in a clinical pathway.	Supports feasibility of AI-supported referral/screening, not unsupervised counseling.
Keskinbora and Guven [3]	Broad review of AI in ophthalmology.	Supports rationale for studying patient-facing AI in eye care.
Costin et al. [4]	Reviewed advantages and limitations of ophthalmic AI.	Aligns with cautious conclusion and need for validation.
Chalasani et al. [5]	Discussed structured evaluation of AI in healthcare practice.	Supports standardized assessment of counseling tools.
Zhang et al. [45]	Systematic review of LLMs in ophthalmology.	Confirms rapid growth and heterogeneity; this review is narrower and patient-facing.
Agnihotri et al. [46]	Review of LLM publications in ophthalmology journals.	Supports need for clear model, prompt, and outcome reporting.
Artiaga et al. [47]	Scoping review of LLM utility for clinicians, patients, researchers, and educators.	This review adds focus on counseling and education outcomes.
Li et al. [48]	Systematic review of LLMs in primary care ophthalmic education.	Closely aligned with education focus; this review also includes referral/follow-up support.
Andrew and Rodenko [49]	Narrative review of LLMs and eye education.	Supports emphasis on readability, health literacy, and safety.
See et al. [50]	Scoping review of ophthalmology LLM use-cases.	Supports governance and implementation safeguards.
Riazi Esfahani et al. [38]	Assessed ChatGPT accuracy and readability for common ophthalmology questions.	Direct evidence that accuracy and readability must be assessed together.
Dal et al. [39]	Assessed Turkish glaucoma patient education materials.	Shows need for multilingual chronic-disease counseling validation.
Yang et al. [40]	Assessed pediatric ophthalmology surgery education materials.	Shows AI material is not always superior to society material.
Bondok et al. [41]	Compared ophthalmologist and chatbot responses to patient questions.	Supports supervised drafting, not autonomous replacement.

Table 5. Cont.

Reference	Main Contribution	Comparison with This Review
Lin et al. [42]	Evaluated glaucoma counseling in Traditional Chinese.	Shows language-specific validation is important.
Wu et al. [43]	Compared ChatGPT-4o handouts with AAO materials.	Supports use as content-generation aid with expert review.
Hamzeh et al. [44]	Evaluated ChatGPT-4o accuracy and readability for patient questions.	Reinforces need for both clinical and readability review.

4.3. Accuracy, Completeness, and Clinical Appropriateness

Accuracy was the most visible outcome across the included studies, but it was measured in different ways. Some studies used expert ratings, some compared answers with professional guidelines or society materials, and others used study-specific rubrics. This makes direct pooling inappropriate, but the pattern is clinically useful: AI systems can produce correct and coherent responses in many eye-care topics, yet errors, omissions, overgeneralized advice, and variation between models remain important concerns.

This finding agrees with broader LLM reviews in ophthalmology, which describe encouraging performance but emphasize that evaluation frameworks are not standardized [45–50]. For patient counseling, variability is especially important because incomplete or overconfident answers may affect when patients seek care. An AI answer about cataract surgery, glaucoma drops, diabetic retinopathy follow-up, pediatric surgery, or retinal detachment symptoms may appear clear while still missing individualized risk factors or urgent referral criteria. Therefore, clinical appropriateness should be judged not only by factual correctness but also by whether the response is safe for patient behavior.

The review also suggests that model comparisons should be interpreted carefully. No model family consistently outperformed all others across ophthalmic tasks. ChatGPT-family models, including newer GPT-4o-based evaluations, often produced clinically useful patient-facing responses, but studies also reported concerns about length, readability, incompleteness, or missing safety warnings. Gemini/Bard responses were generally comparable in some tasks but sometimes scored lower than ChatGPT-family systems depending on language, prompt, and scoring rubric. Disease-specific and multimodal systems may offer advantages when they are designed for a narrow task, such as myopia consultation, fundus-fluorescein angiography question answering, or ophthalmology-specific patient inquiries, but these advantages should not be generalized beyond the validated task and dataset. Therefore, model choice for clinical education should be guided by the clinical topic, language, patient population, prompting method, comparator, and requirement for clinician review rather than by model name alone.

Readability and patient-education quality should be interpreted alongside accuracy. Several included studies showed that AI can simplify ophthalmic terminology and generate structured educational materials, but responses may still be too long, too technical, or above recommended patient education levels. Therefore, AI-generated counseling should be judged not only by factual correctness but also by completeness, understandability, actionability, and safety for patient behavior.

Multilingual and culturally appropriate communication also require attention. Studies in Turkish and Traditional Chinese settings show that AI evaluation cannot be limited to English-language prompts [39,42]. Eye-care counseling often occurs in multilingual contexts, and literal translation may not preserve clinical nuance, urgency, or health-literacy appropriateness. Future research should test AI-generated materials in the languages and cultural contexts in which patients will actually use them.

4.4. Patient-Centered Outcomes and Real-World Implementation

A major weakness of the current evidence is the limited measurement of patient-centered outcomes. Many studies evaluated AI outputs rather than patient outcomes. As a result, the literature can tell us whether expert reviewers considered an answer accurate or readable, but it often cannot tell us whether patients understood the answer, trusted it appropriately, followed instructions, attended follow-up, avoided unnecessary anxiety, or avoided delayed care.

This limitation is not unique to the present review; recent ophthalmology LLM reviews also emphasize that real-world evaluation remains underdeveloped [45–50]. For optometry and eye-care practice, this gap is clinically significant. Counseling success is not measured only by the content of the message; it is measured by whether the patient can use the information safely. A future evidence base should therefore include prospective patient studies, pragmatic clinical evaluations, and outcomes such as understanding, recall, satisfaction, trust calibration, adherence, follow-up attendance, escalation of urgent symptoms, and clinician workload.

The role of AI in practice should therefore be framed as supportive rather than substitutive. Evidence-based conclusions from the included studies support the potential of AI to draft, simplify, or structure patient-facing educational content under supervision. By contrast, recommendations for governance, workflow integration, autonomous deployment, adherence support, and long-term follow-up remain conceptual or implementation-oriented because prospective patient studies and real-world clinical deployment data remain limited. AI may help clinicians draft educational explanations, generate first versions of patient handouts, simplify jargon, prepare multilingual materials, or provide structured answers that clinicians can review, but the present evidence does not support unsupervised use as a substitute for clinician-led counseling.

4.5. Safety, Governance, and Clinical Accountability

Safety was inconsistently reported across the included studies, even though it is one of the most important issues for patient-facing AI. Potential safety concerns include misinformation, incomplete warnings, inappropriate reassurance, failure to recommend urgent referral, overconfident statements, hallucinated details, and advice that is not individualized to patient history. In eye care, these concerns are clinically important because delayed assessment may lead to avoidable vision loss in conditions such as retinal detachment, acute glaucoma, severe ocular infection, trauma, or proliferative diabetic retinopathy.

The discussion in broader AI reviews is therefore directly relevant to this manuscript. Previous reviews emphasize that clinical AI requires governance, transparency, accountability, monitoring, and human oversight [3,4,45–50]. For patient education, this means that AI output should be reviewed before being integrated into clinical workflows, especially when advice involves medications, surgery, urgent symptoms, pediatric care, or chronic disease follow-up. Clinics should also clarify to patients that AI-generated information is educational and does not replace personalized assessment by a qualified eye-care professional.

Future implementation should include safeguards such as approved prompt templates, clinician review, clear referral warnings, reading-level targets, version tracking, documentation of model updates, reporting of unsafe outputs, and periodic revalidation. These safeguards are necessary because AI models change over time, and findings from one version may not remain valid after updates or changes in deployment conditions.

4.6. Strengths and Limitations

This review has several strengths. It focuses on a clinically important and rapidly developing area: AI-supported patient counseling, education, and management-related communication in optometry and related eye-care settings. It separates patient-facing communication studies from purely technical image-classification or segmentation studies, which makes the conclusions more relevant to counseling practice. It also synthesizes multiple outcome domains, including accuracy, readability, completeness, appropriateness, safety, usefulness, and patient-centered outcomes when available.

The review also has limitations. First, the included studies were highly heterogeneous in AI model, prompt design, disease area, language, comparator, scoring rubric, and outcome definition. This limited direct comparison between studies and prevented quantitative pooling. Second, many studies were based on simulated prompts, AI-generated outputs, or expert review rather than real patient encounters. Third, reporting of participant characteristics, setting, model version, prompt wording, and evaluator background was incomplete in several studies. Fourth, patient-centered outcomes such as satisfaction, trust, usability, adherence, follow-up behavior, and clinical impact were not consistently measured. Finally, because AI tools evolve rapidly, results may become outdated when models are updated. Although methodological quality was appraised using JBI tools, the diversity of study designs and outcome measures meant that appraisal findings were interpreted narratively rather than used to exclude studies or produce a pooled quality score.

4.7. Implications for Future Research

Future studies should move from output-only evaluations toward prospective and real-world designs. Studies should test AI tools with real patients, caregivers, and clinicians in realistic counseling scenarios. They should use standardized rubrics for accuracy, completeness, appropriateness, readability, actionability, safety, and potential harm. They should also report model version, query date, prompt wording, language, clinical context, reference standard, evaluator expertise, and whether responses were regenerated or edited. Future evaluations should also assess accessibility for low-vision users, including screen-reader compatibility, readable formatting, adequate visual contrast, plain-language structure, and the suitability of AI-generated patient education materials for patients with visual impairment.

Patient-centered outcomes should become a priority. Future research should measure whether AI improves patient understanding, reduces confusion, supports adherence, improves follow-up attendance, increases appropriate care-seeking, or reduces clinician workload without compromising safety. Multilingual studies and low-health-literacy evaluations are especially important for equitable implementation. In the meantime, AI in optometry and related eye-care settings should be used as a clinician-supervised support tool for education and communication rather than an autonomous replacement for professional counseling.

5. Conclusions

This systematic review included 39 studies focused on patient-facing AI applications for counseling, education, communication, referral/follow-up, or management support in optometry and related eye-care settings. AI-powered systems, particularly LLMs and chatbot-based tools, show meaningful potential as supervised adjuncts for patient education and management-related communication. However, the current evidence is derived primarily from expert evaluations, simulated prompts, AI-generated outputs, content comparisons, and limited clinical pathway studies rather than from prospective real-world deployment with patient-centered outcomes. Therefore, conclusions about accuracy, read-

ability, clinical appropriateness, and usefulness should be interpreted cautiously, and recommendations for implementation should be viewed as future-facing until validated in real patients, diverse languages, and routine clinical settings. Before routine adoption, studies should use standardized evaluation frameworks, report model and prompt details transparently, monitor safety and potential harm, and measure patient understanding, trust, adherence, follow-up behavior, and clinical outcomes.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/healthcare14152270/s1>, File S1. PRISMA 2020 Checklist and Search Strategy. Table S1. Final Included Studies and DOI/Stable Identifier Audit ($n = 39$). Table S2. Study-by-study JBI methodological quality appraisal of the 39 included studies. All included studies listed in the Supplementary Materials are represented in the main reference list.

Author Contributions: Conceptualization, M.M.A. and E.M.A.; methodology, A.A.A., M.M.A. and E.M.A.; validation, M.M.A., E.M.A. and Z.A.A.-H.; formal analysis, A.A.A., M.M.A. and E.M.A.; investigation, Z.A.A.-H., A.A., H.A., R.A., M.A., M.M.A. and E.M.A.; data curation, Z.A.A.-H., A.A.A., A.A., H.A., R.A. and M.A.; writing—original draft preparation, Z.A.A.-H., M.M.A. and E.M.A.; writing—review and editing, M.M.A., E.M.A., Z.A.A.-H., A.A.A., M.A., R.A., H.A. and A.A.; supervision, M.M.A. and E.M.A.; project administration, M.M.A., E.M.A., A.A.A. and H.A.; funding acquisition, M.M.A. and E.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was supported by a grant from the “Research Center of the Female Scientific and Medical Colleges”, Deanship of Scientific Research, King Saud University. No specific grant number was provided. The funder had no role in the study design, data collection, data analysis, interpretation of findings, manuscript preparation, or decision to submit the article for publication.

Institutional Review Board Statement: Not applicable. This systematic review used the published literature and did not involve direct collection of data from human participants or identifiable patient information.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting this systematic review are included in the article and Supplementary Materials. Further inquiries can be directed to the corresponding authors.

Acknowledgments: The authors extend their appreciation to the “Research Center of the Female Scientific and Medical Colleges”, Deanship of Scientific Research, King Saud University, for supporting this research project.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gulshan, V.; Peng, L.; Coram, M.; Stumpe, M.C.; Wu, D.; Narayanaswamy, A.; Venugopalan, S.; Widner, K.; Madams, T.; Cuadros, J.; et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* **2016**, *316*, 2402–2410. [[CrossRef](#)] [[PubMed](#)]
2. Abramoff, M.D.; Lavin, P.T.; Birch, M.; Shah, N.; Folk, J.C. Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ Digit. Med.* **2018**, *1*, 39. [[CrossRef](#)] [[PubMed](#)]
3. Keskinbora, K.; Guven, F. Artificial intelligence and ophthalmology. *Turk. J. Ophthalmol.* **2020**, *50*, 37–43. [[CrossRef](#)] [[PubMed](#)]
4. Costin, H.N.; Fira, M.; Goras, L. Artificial intelligence in ophthalmology: Advantages and limits. *Appl. Sci.* **2025**, *15*, 1913. [[CrossRef](#)]
5. Chalasani, S.H.; Syed, J.; Ramesh, M.; Patil, V.; Pramod Kumar, T.M. Artificial intelligence in the field of pharmacy practice: A literature review. *Explor. Res. Clin. Soc. Pharm.* **2023**, *12*, 100346. [[CrossRef](#)] [[PubMed](#)]
6. Young, J.A.; Chang, C.W.; Scales, C.W.; Menon, S.V.; Holy, C.E.; Blackie, C.A. Machine learning methods using artificial intelligence deployed on electronic health record data for identification and referral of at-risk patients from primary care physicians to eye care specialists: Retrospective, case-controlled study. *JMIR AI* **2024**, *3*, e48295. [[CrossRef](#)] [[PubMed](#)]

7. Dihan, Q.A.; Brown, A.D.; Chauhan, M.Z.; Alzein, A.F.; Abdelnaem, S.E.; Kelso, S.D.; Rahal, D.A.; Park, R.; Ashraf, M.; Azzam, A.; et al. Leveraging large language models to improve patient education on dry eye disease. *Eye* **2025**, *39*, 1115–1122. [[CrossRef](#)] [[PubMed](#)]
8. Wei, L.; Mohammed, I.S.K.; Francomacaro, S.; Munir, W.M. Evaluating text-based generative artificial intelligence models for patient information regarding cataract surgery. *J. Cataract. Refract. Surg.* **2024**, *50*, 95–96. [[CrossRef](#)] [[PubMed](#)]
9. Wu, Y.; Chen, X.; Zhang, W.; Liu, S.; Sum, W.M.R.; Wu, X.; Shang, X.; Kee, C.-S.; He, M.; Shi, D. ChatMyopia: An AI agent for myopia-related consultation in primary eye care settings. *iScience* **2025**, *28*, 113768. [[CrossRef](#)] [[PubMed](#)]
10. Yilmaz, İ.E.; Berhuni, M.; Özer Özcan, Z.; Doğan, L. Chatbots talk strabismus: Can AI become the new patient educator? *Int. J. Med. Inform.* **2024**, *191*, 105592. [[CrossRef](#)] [[PubMed](#)]
11. Wu, H.; Su, Z.; Pan, X.; Shao, A.; Xu, Y.; Wang, Y.; Jin, K.; Ye, J. Enhancing diabetic retinopathy query responses: Assessing large language model in ophthalmology. *Br. J. Ophthalmol.* **2025**, *109*, 1272–1278. [[CrossRef](#)] [[PubMed](#)]
12. Wu, D.; Li, Y.; Zhang, H.; Yang, X.; Mao, Y.; Chen, B.; Feng, Y.; Chen, L.; Zou, X.; Nie, Y.; et al. An artificial intelligence platform for the screening and managing of strabismus. *Eye* **2024**, *38*, 3101–3107. [[CrossRef](#)] [[PubMed](#)]
13. Armadá-Maresca, F.; Capote-Díaz, P.; Caminos-Melguizo, A.; Baoud-Ould-Haddi, I. Clinical performance of an interactive platform based on artificial intelligence in ophthalmology: Experience in a third-level reference center. *Front. Med.* **2025**, *12*, 1593556. [[CrossRef](#)] [[PubMed](#)]
14. Strzalkowski, P.; Strzalkowska, A.; Chhablani, J.; Pfau, K.; Errera, M.-H.; Roth, M.; Schaub, F.; Bechrakis, N.E.; Hoerauf, H.; Reiter, C.; et al. 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. Retin. Vitro.* **2024**, *10*, 61. [[CrossRef](#)] [[PubMed](#)]
15. Tailor, P.D.; D'Souza, H.S.; Castillejo Becerra, C.M.; Dahl, H.M.; Patel, N.R.; Kaplan, T.M.; Kohli, D.; Bothun, E.D.; Mohny, B.G.; Tooley, A.A.; et al. Evaluation of AI summaries on interdisciplinary understanding of ophthalmology notes. *JAMA Ophthalmol.* **2025**, *143*, 410–419. [[CrossRef](#)] [[PubMed](#)]
16. Raghu, K.; Tamilselvi, S.; Devishamani, C.S.; Suchetha, M.; Rajalakshmi, R.; Raman, R. The utility of ChatGPT in diabetic retinopathy risk assessment: A comparative study with clinical diagnosis. *Clin. Ophthalmol.* **2023**, *17*, 4021–4031. [[CrossRef](#)] [[PubMed](#)]
17. Wolf, R.M.; Channa, R.; Liu, T.Y.A.; Zehra, A.; Bromberger, L.; Patel, D.; Ananthakrishnan, A.; Brown, E.A.; Prichett, L.; Lehmann, H.P.; et al. Autonomous artificial intelligence increases screening and follow-up for diabetic retinopathy in youth: The ACCESS randomized control trial. *Nat. Commun.* **2024**, *15*, 421. [[CrossRef](#)] [[PubMed](#)]
18. Dow, E.R.; Khan, N.C.; Chen, K.M.; Mishra, K.; Perera, C.; Narala, R.; Basina, M.; Dang, J.; Kim, M.; Levine, M.; et al. AI-human hybrid workflow enhances teleophthalmology for the detection of diabetic retinopathy. *Ophthalmol. Sci.* **2023**, *3*, 100330. [[CrossRef](#)] [[PubMed](#)]
19. Shruthee, S.G.; Vellingiri, D.; Rajagopal, K.; Shanmugaiah, P. Diagnostic accuracy of an artificial intelligence based mHealth intervention for cataract detection: A multi-center prospective study in Tamil Nadu, India. *Natl. J. Community Med.* **2025**, *16*, 887–894. [[CrossRef](#)]
20. Gupta, A.S.; Sulewski, M.E.; Armenti, S.T. Performance of ChatGPT in cataract surgery counseling. *J. Cataract. Refract. Surg.* **2024**, *50*, 424–425. [[CrossRef](#)] [[PubMed](#)]
21. Yigen İritaş, S.; İritaş, İ. AI as a patient educator: Evaluating ChatGPT's role in disseminating information on herpes zoster ophthalmicus. *Eur. Res. J.* **2025**, *11*, 769–775. [[CrossRef](#)]
22. Ma, R.; Cheng, Q.; Yao, J.; Peng, Z.; Yan, M.; Lu, J.; Liao, J.; Tian, L.; Shu, W.; Zhang, Y.; et al. Multimodal machine learning enables AI chatbot to diagnose ophthalmic diseases and provide high-quality medical responses. *NPJ Digit. Med.* **2025**, *8*, 64. [[CrossRef](#)] [[PubMed](#)]
23. Mathenge, W.; Whitestone, N.; Nkurikiye, J.; Patnaik, J.L.; Piyasena, P.; Uwaliraye, P.; Lanouette, G.; Kahook, M.Y.; Cherwek, D.H.; Congdon, N.; et al. Impact of artificial intelligence assessment of diabetic retinopathy on referral service uptake in a low-resource setting: The RAIDERS randomized trial. *Ophthalmol. Sci.* **2022**, *2*, 100168. [[CrossRef](#)] [[PubMed](#)]
24. Mereddy, S.; Polaki, S.; Aswini Kumar, M.; Mohanababu, C.; Prathibha Malini, K.; Sujatha, D. AI and IoT-integrated smart devices for enhanced vision screening and diagnosis in eye care. In *2024 2nd International Conference on Signal Processing, Communication, Power and Embedded Systems (SCOPES)*; IEEE: Piscataway, NJ, USA, 2024. [[CrossRef](#)]
25. Reiter, G.S.; Lachinov, D.; Bühl, W.; Weigert, G.; Grechenig, C.; Mai, J.; Bogunović, H.; Schmidt-Erfurth, U. A novel management challenge in age-related macular degeneration: Artificial intelligence and expert prediction of geographic atrophy. *Ophthalmol. Retin.* **2025**, *9*, 421–430. [[CrossRef](#)] [[PubMed](#)]
26. Karataş, E.; Durmaz Engin, C. From accuracy to comprehensibility: Evaluating large language models for myopia patient queries. *Health Policy Technol.* **2025**, *14*, 101073. [[CrossRef](#)]
27. Dow, E.R.; Chen, K.M.; Zhao, C.S.; Knapp, A.; Phadke, A.; Weng, K.; Do, D.V.; Mahajan, V.B.; Mruthyunjaya, P.; Leng, T.; et al. Artificial intelligence improves patient follow-up in a diabetic retinopathy screening program. *Clin. Ophthalmol.* **2023**, *17*, 3323–3330. [[CrossRef](#)] [[PubMed](#)]

28. Al-Sharif, E.M.; Penteado, R.C.; Atwan, M.; El Jalbout, N.D.; Topilow, N.J.; Shoji, M.K.; Kikkawa, D.O.; Liu, C.Y.M.; Korn, B.S.M. Evaluating the accuracy of ChatGPT and Google BARD in fielding oculoplastic patient queries: A comparative study on artificial versus human intelligence. *Ophthalmic Plast. Reconstr. Surg.* **2024**, *40*, e95–e100. [\[CrossRef\]](#) [\[PubMed\]](#)
29. Kufta, A.Y.; Djalilian, A. Enhancing patient education with AI: A readability analysis of AI-generated versus American Academy of Ophthalmology online patient education materials. *J. Clin. Med.* **2025**, *14*, 6968. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Balas, M.; Kaplan, A.J.; Esmail, K.; Saleh, S.; Sharma, R.A.; Yan, P.; Arjmand, P. Translating ophthalmic medical jargon with artificial intelligence: A comparative comprehension study. *Can. J. Ophthalmol.* **2025**, *60*, e602–e609. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Cao, Y.; Lu, W.; Shi, R.; Liu, F.; Liu, S.; Xu, X.; Yang, J.; Rong, G.; Xin, C.; Zhou, X.; et al. Performance of popular large language models in glaucoma patient education: A randomized controlled study. *Adv. Ophthalmol. Pract. Res.* **2025**, *5*, 88–94. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Li, G.; Lin, M.X.; Cui, D.; Mathews, P.M.; Akpek, E.K. Enhancing online cataract surgery patient education materials through artificial intelligence. *Can. J. Ophthalmol.* **2025**, *60*, e197–e202. [\[CrossRef\]](#) [\[PubMed\]](#)
33. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Lee, G.G.; Goodman, D.; Chang, T.C.P. Impact of demographic modifiers on readability of myopia education materials generated by large language models. *Clin. Ophthalmol.* **2024**, *18*, 3591–3604. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Chen, X.; Zhao, Z.; Zhang, W.; Xu, P.; Wu, Y.; Xu, M.; Gao, L.; Li, Y.; Shang, X.; Shi, D.; et al. EyeGPT for patient inquiries and medical education: Development and validation of an ophthalmology large language model. *J. Med. Internet Res.* **2024**, *26*, e60063. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Anguita, R.; Downie, C.; Ferro Desideri, L.; Sagoo, M.S. Assessing large language models' accuracy in providing patient support for choroidal melanoma. *Eye* **2024**, *38*, 3092–3098. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Macri, C.Z.; Bacchi, S.; Wong, W.; Baranage, D.; Sivagurunathan, P.D.; Chan, W.O. A pilot survey of patient perspectives on an artificial intelligence-generated presenter in a patient information video about face-down positioning after vitreoretinal surgery. *Ophthalmic Res.* **2024**, *67*, 567–572. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Riaz Esfahani, P.; Ward, J.; Yong, A.; Nguyen, T.B.; Reddy, A.J.; Sobhani, S.; Chen, D.; Akanda, M.; Sheikh, S. Evaluating ChatGPT's accuracy and readability in responding to common ophthalmology questions. *Cureus* **2025**, *17*, e87920. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Dal, A.; Erdağ, M.; Dikme, B.; Kutluksaman, B. Use of large language models in Turkish information materials for glaucoma patient education: Evaluation of readability, accuracy and comprehensiveness. *Eur. Eye Res.* **2026**, *6*, 60–69. [\[CrossRef\]](#)
40. Yang, A.; Reid, M.; Nguyen, A.M.; Nallasamy, S.; Velez, F.G.; de Alba Campomanes, A.G.; Chang, M.Y. Quality and readability of patient educational materials generated by ChatGPT-4o for pediatric ophthalmologic surgeries. *J. Pediatr. Ophthalmol. Strabismus* **2025**, *62*, 421–427. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Bondok, M.; Selvakumar, R.; Law, C.; Ing, E.B.; Bakshi, N.K.; Felfeli, T. Comparing ophthalmologist and artificial intelligence chatbot responses to patient questions. *Clin. Ophthalmol.* **2025**, *19*, 4293–4300. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Lin, C.W.; Lin, Y.C.; Chen, C.Y.; Wang, J.J.; Lin, H.Y.; Lai, I.C. Evaluating large language models for glaucoma counseling: A pilot study of ChatGPT and Google Gemini responses in Traditional Chinese. *Taiwan J. Ophthalmol.* **2026**, *16*, 141–145. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Wu, L.L.; Hong, A.T.; Davuluru, S.S.; Moreno-Diaz, E.A.; Gutiérrez-Hernández, J.C.; Martínez-Camarillo, J.C.; Xu, B.Y.; Zhang-Nunes, S. Utility of ChatGPT-4o in creating patient handouts in ophthalmology: A comparison with American Academy of Ophthalmology educational materials. *Transl. Vis. Sci. Technol.* **2026**, *15*, 14. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Hamzeh, N.; Lidder, A.K.; Feder, R.S.; Sarmiento, E.A.; Mirza, R.G.; Thau, A.J.; Tanna, A.P. Accuracy and readability of Chat Generative Pre-trained Transformer-4 Omni in answering ophthalmology patient questions. *Ophthalmol. Sci.* **2026**, *6*, 101007. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Zhang, Z.; Zhang, H.; Pan, Z.; Bi, Z.; Wan, Y.; Song, X.; Fan, X. Evaluating large language models in ophthalmology: Systematic review. *J. Med. Internet Res.* **2025**, *27*, e76947. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Agnihotri, A.P.; Nagel, I.D.; Artiaga, J.C.M.; Guevarra, M.C.B.; Sosuan, G.M.N.; Kalaw, F.G.P. Large language models in ophthalmology: A review of publications from top ophthalmology journals. *Ophthalmol. Sci.* **2025**, *5*, 100681. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Artiaga, J.C.M.; Guevarra, M.C.B.; Sosuan, G.M.N.; Agnihotri, A.P.; Nagel, I.D.; Kalaw, F.G.P. Large language models in ophthalmology: A scoping review on their utility for clinicians, researchers, patients, and educators. *Eye* **2025**, *39*, 2752–2761. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Li, S.; Wang, X.; Chen, Y.; Tian, M.; Lin, P.; Lai, M.; Jiang, L. Large language models for primary care ophthalmic education: A systematic review. *Front. Med.* **2026**, *13*, 1810098. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Andrew, A.; Rodenko, S. Large language models and eye education: A narrative review. *Oman J. Ophthalmol.* **2026**, *19*, 17–27. [\[CrossRef\]](#) [\[PubMed\]](#)

50. See, Y.K.C.; Lim, K.S.A.; Au, W.Y.; Chia, S.Y.C.; Fan, X.; Li, Z.K. The use of large language models in ophthalmology: A scoping review on current use-cases and considerations for future works in this field. *Big Data Cogn. Comput.* **2025**, *9*, 151. [[CrossRef](#)]
51. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)] [[PubMed](#)]
52. Aromataris, E.; Munn, Z. (Eds.) *JBIR Manual for Evidence Synthesis*; JBI: Adelaide, Australia, 2020. Available online: <https://jbi-global-wiki.refined.site/space/MANUAL>. [[CrossRef](#)]
53. Rethlefsen, M.L.; Kirtley, S.; Waffenschmidt, S.; Ayala, A.P.; Moher, D.; Page, M.J.; Koffel, J.B.; PRISMA-S Group. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst. Rev.* **2021**, *10*, 39. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.