

# Comparing the Accuracy and Readability of Glaucoma-related Question Responses and Educational Materials by Google and ChatGPT

Samuel A Cohen<sup>1</sup>, Ann C Fisher<sup>2</sup>, Benjamin Y Xu<sup>3</sup>, Brian J Song<sup>4</sup>

Received on: 15 June 2024; Accepted on: 17 September 2024; Published on: 29 October 2024

## ABSTRACT

**Aim and background:** Patients are increasingly turning to the internet to learn more about their ocular disease. In this study, we sought (1) to compare the accuracy and readability of Google and ChatGPT responses to patients' glaucoma-related frequently asked questions (FAQs) and (2) to evaluate ChatGPT's capacity to improve glaucoma patient education materials by accurately reducing the grade level at which they are written.

**Materials and methods:** We executed a Google search to identify the three most common FAQs related to 10 search terms associated with glaucoma diagnosis and treatment. Each of the 30 FAQs was inputted into both Google and ChatGPT and responses were recorded. The accuracy of responses was evaluated by three glaucoma specialists while readability was assessed using five validated readability indices. Subsequently, ChatGPT was instructed to generate patient education materials at specific reading levels to explain seven glaucoma procedures. The accuracy and readability of procedural explanations were measured.

**Results:** ChatGPT responses to glaucoma FAQs were significantly more accurate than Google responses (97 vs 77% accuracy, respectively,  $p < 0.001$ ). ChatGPT responses were also written at a significantly higher reading level (grade 14.3 vs 9.4, respectively,  $p < 0.001$ ). When instructed to revise glaucoma procedural explanations to improve understandability, ChatGPT reduced the average reading level of educational materials from grade 16.6 (college level) to grade 9.4 (high school level) ( $p < 0.001$ ) without reducing the accuracy of procedural explanations.

**Conclusion:** ChatGPT is more accurate than Google search when responding to glaucoma patient FAQs. ChatGPT successfully reduced the reading level of glaucoma procedural explanations without sacrificing accuracy, with implications for the future of customized patient education for patients with varying health literacy.

**Clinical significance:** Our study demonstrates the utility of ChatGPT for patients seeking information about glaucoma and for physicians when creating unique patient education materials at reading levels that optimize understanding by patients. An enhanced patient understanding of glaucoma may lead to informed decision-making and improve treatment compliance.

**Keywords:** Artificial intelligence, ChatGPT, Glaucoma, Google, Patient education.

*Journal of Current Glaucoma Practice* (2024): 10.5005/jp-journals-10078-1448

## INTRODUCTION

As access to reliable internet across the United States has continued to increase, so, too, has the propensity for patients to use the internet to seek information about their health.<sup>1,2</sup> When searching for health information, patients often turn to search engines such as Google, which accounts for >90% of total search engine queries and approximately one billion health-related searches worldwide each day.<sup>3</sup> Despite "Dr Google's" popularity, several studies have documented that health information provided by Google relating to a variety of ophthalmology subspecialties including retina, glaucoma, and oculoplastics is often misleading and can be difficult to understand for patients without advanced higher education degrees.<sup>4-9</sup>

The rising popularity of artificial intelligence (AI) large language models (LLMs) offers an additional online modality that patients can use to learn more about their health. One notable LLM that has surged in popularity in the United States is ChatGPT (Open AI), which gained >1 million users in just 5 days after launching in 2022 and continues to amass >1 billion monthly website queries.<sup>10,11</sup> Although ChatGPT has been shown to be useful in medicine *via* its ability to assist in writing progress notes and discharge summaries, evidence regarding its ability to directly respond to patient questions is mixed.<sup>12,13</sup> Within ophthalmology, one study claimed that ChatGPT responses to patient questions are often inaccurate and potentially

<sup>1</sup>Department of Ophthalmology, UCLA Stein Eye Institute, Los Angeles, California, United States

<sup>2</sup>Department of Ophthalmology Byers Eye Institute at Stanford, Stanford, California, United States

<sup>3,4</sup>Department of Ophthalmology, USC Roski Eye Institute, Los Angeles, California, United States

**Corresponding Author:** Brian J Song, Department of Ophthalmology, USC Roski Eye Institute, Los Angeles, California, United States, Phone: +1 3234426335, e-mail: brian.song@med.usc.edu

**How to cite this article:** Cohen SA, Fisher AC, Xu BY, *et al.* Comparing the Accuracy and Readability of Glaucoma-related Question Responses and Educational Materials by Google and ChatGPT. *J Curr Glaucoma Pract* 2024;18(3):110-116.

**Source of support:** Nil

**Conflict of interest:** None

harmful; however, several others have found ChatGPT responses to patient queries to be accurate and preferable to those found on other patient education websites.<sup>14-16</sup>

While the ability of popular online search modalities such as Google and ChatGPT to respond to patients' frequently asked questions (FAQs) has been studied in various ophthalmology subspecialties including

retina, cornea, and oculoplastics, there is a paucity of information on how these tools respond to glaucoma patient questions.<sup>14,15,17-19</sup> Additionally, there is limited information to date regarding the ability of AI LLMs like ChatGPT to revise existing glaucoma patient education materials to make them more understandable without sacrificing the accuracy of the information presented.

This study has two main objectives. First, we compare the accuracy and readability of Google and ChatGPT responses to common glaucoma patient FAQs. Second, we evaluate whether ChatGPT is capable of improving glaucoma patient education materials by reducing the grade level at which they are written while maintaining their accuracy—with the goal of providing custom education for patients with varying health literacy levels.

## MATERIALS AND METHODS

This study was approved by the Stanford University Institutional Review Board (IRB).

### Frequently Asked Questions Selection and Responses

For all Google searches used in data collection, we utilized a clean-installed Google Chrome (Menlo Park, California, United States) browser on Incognito Mode. We disabled sponsored results and location filters to eliminate bias from prior searches and targeted geographic search results. We selected the 10 search terms in this study based on a recently published manuscript evaluating glaucoma patient education materials.<sup>4</sup> The search terms were the following: “glaucoma,” “open-angle glaucoma,” “angle-closure glaucoma,” “high eye pressure,” “glaucoma surgery,” “glaucoma eye drops,” “minimally invasive glaucoma surgery,” “trabeculectomy,” “tube shunt,” and “glaucoma treatments.”

For all 10 search terms, a Google search was executed and the first three FAQs associated with each respective search term were noted. These 30 FAQs were subsequently entered into both the Google search engine and the ChatGPT V 3.5 LLM and responses generated by both modalities were recorded. For Google responses, the direct quote provided by Google in response to the FAQ as well as the associated website from which the text was pulled to generate the response were noted. Google responses were solely generated from content written on websites that populated after Google search and were not generated by AI. For ChatGPT LLM responses, the direct text provided by the AI LLM was recorded.

### Readability Analysis

The readability of responses provided by Google and ChatGPT to patient FAQs was measured using five validated readability assessments: Flesch Reading Ease (FRE), Gunning Fog Index (GFI), Flesch–Kincaid Grade Level (FKGL), Simple Measure of Gobbledygook (SMOG), and Coleman Liau Index (CLI). While the FRE scale measures readability by generating a score from 0 (very difficult) to 100 (very easy), all other remaining indices generate a “grade level” at which a block of text is written. For example, a score of 9 indicates the text was written at a ninth grade (high school) reading level. The GFI, FKGL, SMOG, and CLI indices were used to generate an average grade level for each block of text.

### JAMA Accountability Analysis

The accountability of the 30 websites utilized by Google to generate responses to patient FAQs was assessed on a 0–4 point scale using JAMA benchmarks. Each website received one point for each of the following JAMA accountability metrics: (1) listing all authors

and their relevant credentials, (2) listing references, (3) providing disclosures, and (4) providing date of last update.

### ChatGPT Procedure Explanations

To assess the ability of ChatGPT to help patients better understand glaucoma procedures, ChatGPT was instructed to generate patient education materials for seven common glaucoma procedures: trabeculectomy, tube shunt, selective laser trabeculoplasty, laser peripheral iridotomy, canaloplasty, goniotomy, and minimally invasive glaucoma surgery. The phrase, “ChatGPT, please explain what happens in a (procedure name)” was entered into the ChatGPT tool. The responses provided were recorded and evaluated for readability using five validated readability indices. To assess ChatGPT’s ability to generate patient education materials at a specific reading level, the tool was then asked to generate procedure-specific education materials at a seventh grade reading level. The phrase, “ChatGPT, please explain what happens in a (procedure name) at a seventh grade reading level” was entered into the ChatGPT tool. ChatGPT responses were recorded and readability was measured.

### Expert Panel Evaluation: Frequently Asked Questions

Responses to all 30 FAQs generated by both Google and AI were independently reviewed by a panel of three fellowship-trained glaucoma surgeons employed by two different academic institutions for several different criteria. Responses to patient queries provided by Google and AI were listed side by side in a randomized order and three questions were asked. First, experts were asked to select the better (more accurate and comprehensible) response to the patient’s question. They were then asked to identify the response which was generated by an AI LLM. Finally, to evaluate accuracy, they were asked if either of the responses contained inappropriate or inaccurate information.

### Expert Panel Evaluation: Procedures

For all seven glaucoma procedures, the explanations provided by ChatGPT at various reading levels (unspecified and seventh grade level requests) were listed side by side in a randomized order. Experts were first asked to choose which block of text they would select for informational pamphlets designed to better explain the procedure. They were then asked to identify which block of text was written at a lower reading level. Finally, they were asked whether the responses contained inappropriate or inaccurate information.

### Statistical Analysis

Readability of Google and ChatGPT responses to patient FAQs as well as ChatGPT-generated procedural explanations (at various specified reading levels) were compared *via* two-sample *t*-tests. Two-sided  $\chi^2$  tests of independence and two-sided *z*-tests were used to assess for associations between categorical variables as appropriate. A *p*-value < 0.05 was considered statistically significant. Data was analyzed in R version 4.3.2.

## RESULTS

### Glaucoma Frequently Asked Patient Questions

Table 1 displays the 30 FAQs that populated after executing a Google search for 10 search terms related to glaucoma diagnosis and treatment. Patient queries were both qualitative and quantitative.

**Readability of Responses to Patient Frequently Asked Questions: Google vs ChatGPT**

ChatGPT responses to patient FAQs were written at a significantly higher average grade level than Google responses (14.3 vs 9.4, respectively,  $p < 0.001$ ) (Table 2). ChatGPT responses averaged a lower (more difficult to comprehend) FRE than Google responses (35.2 vs 53.0, respectively) ( $p < 0.001$ ).

**JAMA Accountability of Web Pages Providing Answers to Frequently Asked Questions on Google**

The majority of web pages providing answers to Google FAQs related to glaucoma diagnosis and treatments were from private practices and educational institutions. The average JAMA accountability score of the 30 web pages analyzed was 1.53/4. The most frequent JAMA accountability metrics on web pages analyzed were date of last update (70.0%), author list (40.0%), and references (40.0%).

**Readability of ChatGPT Generated Patient Education Materials for Common Glaucoma Procedures**

The readability metrics of ChatGPT-generated patient education materials for common glaucoma procedures are recorded in

(Tables 2 and 3). Without specifying a grade level at which patient education materials should be written, the average grade level of education materials produced was 16.6, with an FRE of 33.1 (“very difficult”). When instructed to generate education materials at a seventh grade reading level, readability significantly improved to an average grade level of 9.4 and an FRE of 67.4 (“standard”) ( $p < 0.001$ ).

**Expert Panel Evaluation: Google and ChatGPT Responses to Frequently Asked Questions**

When asked to identify the better (more accurate and comprehensible) response to the 30 glaucoma FAQs, ChatGPT responses were selected 75.6% (68/90) of the time by expert reviewers, with Google responses chosen less frequently (14.4%, 13/90). The distribution of how expert reviewers compared AI and Google responses to patient FAQs by topic is displayed in Table 4.

Expert reviewers had varied success when asked to identify which of the provided responses to the 30 FAQs was AI-generated. Approximately, 4.4% of expert responses (4/90) were, “I cannot tell which response was generated by an artificial intelligence large language model.” Among the responses that experts attempted to identify, 39.5% (34/86) were correctly identified as the AI-generated response (Table 5). When examining the success with which each

**Table 1:** Most frequently asked questions related to glaucoma diagnosis and treatment

Category	Frequently asked question 1	Frequently asked question 2	Frequently asked question 3
Glaucoma	What is the main cause of glaucoma?	Who usually gets glaucoma?	How do you fix glaucoma in the eye?
Open-angle glaucoma	How fast does glaucoma spread?	How serious is open-angle glaucoma?	What does vision look like with open-angle glaucoma?
Angle closure glaucoma	How does angle-closure glaucoma present?	What is the most common cause of angle closure glaucoma?	When is it too late to treat glaucoma?
High eye pressure	What causes high pressure in your eyes?	What time of day is eye pressure highest?	What medications increase eye pressure?
Glaucoma surgery	How does glaucoma surgery work?	What kind of surgery do they do for glaucoma?	What is the safest glaucoma surgery?
Glaucoma eye drops	Can you ever stop taking glaucoma drops?	What are common side effects of glaucoma medications?	What is the best tolerated eye drops for glaucoma patients?
Minimally invasive glaucoma surgery	What is minimally invasive glaucoma surgery?	What is the success rate of laser surgery for glaucoma?	What are the benefits of MIGS?
Trabeculectomy	What is the disadvantage of trabeculectomy?	What does the eye look like after trabeculectomy?	Is trabeculectomy a major operation?
Tube shunt	How does a tube shunt work?	Is tube shunt surgery painful?	Where is a tube shunt located?
Glaucoma treatments	Can glaucoma be cured?	What triggers glaucoma attacks?	How often should glaucoma patients be checked?

**Table 2:** Readability of responses to glaucoma FAQs – Google vs. ChatGPT

	Flesch Reading Ease (FRE)	Flesch–Kincaid Grade Level (FKGL)	Gunning Fog Index (GFI)	Simple Measure of Gobbledygook (SMOG)	Coleman Liau Index (CLI)	Average grade level
Google results	53.0	7.8	8.6	7.1	14.2	9.4
ChatGPT	35.2	12.8	16.5	11.9	15.9	14.3

Differences in both FRE and average grade level statistically significant at  $p < 0.001$  for both Table 2 and Table 3

**Table 3:** Readability of glaucoma procedure explanations – unspecified grade level vs seventh grade level requested

	Flesch Reading Ease (FRE)	Flesch–Kincaid Grade Level (FKGL)	Gunning Fog Index (GFI)	Simple Measure of Gobbledygook (SMOG)	Coleman Liau Index (CLI)	Average grade level
Unspecified	33.1	14.2	17.8	12.9	16.1	16.6
Seventh grade level	67.4	8.1	10.7	7.8	11.0	9.4

Differences in both FRE and average grade level statistically significant at  $p < 0.001$  for both Table 2 and Table 3



**Table 4:** Expert review of the better (more accurate and comprehensible) responses to patients' frequently asked questions by topic (three questions for each topic reviewed by all three panelists)

Topic	AI	Google	Both answers are equally appropriate
Glaucoma	4	5	0
Open-angle glaucoma	8	1	0
Angle-closure glaucoma	6	1	2
High eye pressure	8	1	0
Glaucoma surgery	6	1	2
Glaucoma eye drops	7	0	2
Minimally invasive glaucoma surgery	8	1	0
Trabeculectomy	7	0	2
Tube shunt	7	1	1
Glaucoma treatments	7	2	0
Total	68	13	9

AI, artificial intelligence

**Table 5:** Expert panelist ability to correctly identify responses to patients' frequently asked questions generated by AI large language models by topic

Topic	Correctly identify artificial intelligence-generated response	Incorrectly label Google response as artificial intelligence-generated response	I cannot tell which response was generated by an AI large language model
Glaucoma	4	4	1
Open-angle glaucoma	3	6	0
Angle-closure glaucoma	4	5	0
High eye pressure	3	6	0
Glaucoma surgery	3	4	2
Glaucoma eye drops	3	6	0
Minimally invasive glaucoma surgery	2	7	0
Trabeculectomy	5	4	0
Tube shunt	4	4	1
Glaucoma treatments	3	6	0
Total	34	52	4

AI, artificial intelligence

**Table 6:** Reading level of patient education information preferred by glaucoma surgeons if they were creating informational pamphlets for patients by topic

Topic	Unspecified reading level (average grade level 15.3)	Seventh grade reading level requested (average grade level 9.4)	Both equally
Trabeculectomy	1	2	0
Tube shunt	3	0	0
Selective laser trabeculoplasty	1	1	1
Laser peripheral iridotomy	2	1	0
Canaloplasty	1	2	0
Goniotomy	2	1	0
Minimally invasive glaucoma surgery	2	1	0
Total	12	8	1

individual grader was able to identify the AI response, scores varied greatly and were 4/26 correct, 2/30 correct, and 28/30 correct for the three respective reviewers.

When asked to evaluate the accuracy of both AI and Google-generated responses to patient FAQs, most responses did not contain inaccurate or inappropriate information (66/90, 73.3%). Google responses were rated to contain inaccurate or inappropriate information 23% of the time while ChatGPT responses contained inaccurate or inappropriate information 3% of the time ( $p < 0.001$ ).

### Expert Panel Evaluation: ChatGPT Generated Patient Education Materials for Common Glaucoma Procedures

When asked to select which block of text (unspecified grade level vs seventh grade reading level requested) they would incorporate into glaucoma procedure education materials, responses were distributed between the text written at the unspecified grade level (12/21, 57.1%), seventh grade reading level (8/21, 38.1%), and both equally (1/21, 4.8%) (Table 6). Panelists were accurately able



to discern which block of text was written at the lower (more appropriate) reading level for most procedures (16/21, 76.2%). When asked to evaluate whether each block of text contained inaccurate or inappropriate information, “neither” was selected 71% of the time. Responses provided at the unspecified grade level were marked inaccurate on 19.0% (4/21) of expert reviews while responses requested at the seventh grade reading level were marked inaccurate on 28.6% (6/21) of expert reviews ( $p = 0.47$ ).

## DISCUSSION

Our study assessed the ability of two popular online interfaces (Google and ChatGPT) to respond to patient FAQs regarding glaucoma diagnosis and treatment. We also sought to demonstrate the ability of LLMs such as ChatGPT to create customized patient education materials for patients with varying health literacy. Despite difficulty with discerning a human-generated response from an LLM-generated response, glaucoma specialists in our study overwhelmingly preferred ChatGPT responses to patient FAQs when compared with Google responses, rating responses provided by the LLM to be significantly more accurate; however, responses provided by ChatGPT were written at a significantly higher reading level than what is recommended by the American Medical Association (AMA), which could limit understandability among patients. ChatGPT did demonstrate utility in creating customizable education materials for patients with varying health literacy. Patient-facing education materials produced by the LLM were largely accurate, and when instructed, ChatGPT was able to successfully reduce the reading level at which the text was written without compromising the accuracy of educational materials. When utilized properly, LLMs such as ChatGPT have the potential to improve how patients receive medical information and how physicians educate patients. AI could also augment patient education efforts and bridge disparities in access to high-quality medical information, thereby ensuring that patients are well-informed when making eye care decisions.

Prior studies from a variety of medical specialties have shown mixed evidence regarding the use of LLMs such as ChatGPT for medical information purposes.<sup>20–22</sup> Within ophthalmology, evidence is also mixed. One recently published study determined that ChatGPT provided incomplete and inaccurate information regarding various ophthalmic conditions.<sup>16</sup> However, our study supports two prior ophthalmology studies that suggest that ChatGPT may be utilized as an accurate information source for patients seeking to learn more about their eye health.<sup>14,15</sup> Our findings align with studies in bariatric surgery, endocrinology, and cardiology, which found LLMs were capable of providing accurate, reproducible responses to patient queries.<sup>20–22</sup>

An LLM with the potential to provide accurate information to patients has tremendous implications for the future of how patients consume health information. Currently, patients most frequently turn to freely available internet search engines (such as Google) to learn more about their health.<sup>2</sup> While quality information regarding glaucoma diagnosis and treatment from respected organizations such as the American Academy of Ophthalmology (AAO) and American Glaucoma Society (AGS) may populate following Google search, prior research shows that these web pages often populate further down on the first page of search results.<sup>6</sup> Patients often click on the first websites that populate after Google search and are therefore less likely to access the information from these more reputable sources. This was evident in the low JAMA accountability scores for websites used by Google to generate responses to

patient questions in this study, as automated responses to patient queries generated by Google are often pulled from the websites at the top, not the bottom, of search results.<sup>23</sup> The use of an LLM such as ChatGPT can help to alleviate this problem by providing a direct response to a patient query, preventing patients from having to sift through medical misinformation when seeking an answer to their glaucoma questions online.<sup>24</sup> An accurate LLM also has the potential to improve access to basic eye care information for patients who may have difficulty scheduling an appointment with an ophthalmologist. Although chatting with an AI LLM cannot be considered a substitute for an in-person or virtual visit with a practicing ophthalmologist, patients have previously reported turning to eye care forums for advice because they did not have easy access to a local ophthalmologist.<sup>15</sup> Disparities in access to care are even more apparent when considering glaucoma specialist care as compared to comprehensive ophthalmology services.<sup>25,26</sup> For patients facing barriers to care, LLMs that provide preliminary information regarding glaucoma diagnosis and treatment may help to inform decisions regarding future treatment options and can serve as a more accurate information source when compared with traditional Google search.

Despite providing responses to patient questions that were graded as more accurate than Google search, responses provided by ChatGPT were written at a significantly higher grade level, which may impair understanding by patients. The 14th grade reading level (college level) of ChatGPT responses may be difficult to interpret for the 62% of Americans who do not have a college degree.<sup>27</sup> Fortunately, our findings support prior studies that demonstrated ChatGPT has the ability to reduce the reading level at which a block of text is written without compromising accuracy.<sup>14,28</sup> In our study, when instructed to make text more easily understandable, glaucoma procedure explanations were successfully revised from an approximately 17th grade reading level (college level) to a ninth grade reading level (high school level). The ability to produce and/or revise patient education materials to make them more understandable has important clinical implications, as prior research demonstrates that providing customizable patient education materials to patients according to their specific health literacy level can lead to a greater understanding of disease, more informed eye care decisions, and improved clinical outcomes.<sup>29–31</sup> This may be particularly important for a condition such as glaucoma which disproportionately affects patients from racial/ethnic groups that, on average, have demonstrated lower health literacy.<sup>32–34</sup> For these patients, revising patient education materials to a reading level even lower than the sixth grade level recommended by the AMA could help to mitigate the racial/ethnic disparities that are observed in glaucoma patient outcomes.<sup>32–34</sup> In the future, it would be beneficial for LLMs such as ChatGPT to default to providing responses at an appropriate reading level; however, in its current form, patients planning to use LLMs to learn more about their eyes should be instructed to specifically request that their questions are answered at a lower reading level in order to improve the chances that they can fully understand the LLM output.

In addition to improving patient education efforts, our findings suggest AI LLMs have numerous practice management implications. The increasing demands placed on ophthalmologists to complete paperwork and other nonclinical tasks have led to a search for ways to make providers more efficient.<sup>35–37</sup> In addition to using AI to help streamline electronic health record communication, LLMs like ChatGPT may help to compose responses to patient queries, messages, and requests so that providers can shift more of their time

to direct patient care.<sup>38</sup> This may be particularly useful for common questions glaucoma specialists receive regarding medication eye drop dosage and frequency as well as both invasive and noninvasive glaucoma treatment modalities. Although responses provided by ChatGPT to patient queries were largely accurate, they were not perfect. Even if providers are not comfortable using responses to patient queries generated by ChatGPT verbatim, the LLM could still make the glaucoma specialist more efficient by generating an initial response that can be further edited for accuracy. LLMs may also be utilized to interact with the patient prior to their visit to gather an interval history since the patient's last appointment, assess medication compliance, and inquire about questions that the patient would like addressed during the upcoming visit. The time allotted to these tasks, which is ordinarily completed by either the ophthalmologist or an associated staff member, can be reallocated to more nuanced aspects of patient care that cannot be addressed by an AI LLM—such as complex discussions regarding treatment options, disease progression, and goals of care. Furthermore, less time spent on nonclinical tasks can lead to more appointment openings, which may help to lessen the burden of the anticipated ophthalmologist shortage and lower barriers to accessing care.<sup>39</sup>

There are limitations to our study. First, while we evaluated responses to FAQs for 10 search terms related to glaucoma diagnosis and treatment, this is a small percentage of all potential web searches related to glaucoma. However, search terms were selected based on a prior study examining online glaucoma patient education materials as well as a Google Trends query of online public search trends.<sup>4</sup> Next, survey responses regarding the accuracy of ChatGPT and Google responses to patient queries were subjective in nature, and the opinions of the three glaucoma specialists in this study may not be representative of all glaucoma specialists. Furthermore, while we assessed readability of responses to patient queries using five validated readability indices, “readable” and “understandable” are not synonyms and there may be a situation in which a patient finds text that is less “readable” to be more understandable, and vice versa; however, the indices used to assess readability in this study have been utilized in several prior studies and are intended to serve as a general proxy for understandability in the medical context.<sup>6,7,14,40,41</sup> Next, the data used to train ChatGPT is not currently available. It is conceivable that some of the websites that populate after Google search were used to generate at least part of the responses to patient queries provided by ChatGPT; however, the extent to which Google search results impact ChatGPT output is unknown. Additionally, postsurvey analysis revealed graders were proficient at differentiating between Google and AI, but that preconceived biases about which one would be “better” likely led to near-total misclassification by two of the three graders. This supports the notion that further work is required to educate patients and providers about the positive potential of AI. Finally, it is important to note that while our study shows tremendous potential for the use of AI LLMs in ophthalmology, information provided by ChatGPT was not 100% accurate and all education materials produced by the current version of ChatGPT should be verified by an ophthalmologist for accuracy and safety.

## CONCLUSION

In conclusion, glaucoma specialists rated ChatGPT responses to glaucoma FAQs to be significantly more accurate than Google responses. Additionally, we demonstrate the utility of ChatGPT in creating customizable patient education materials for patients with

differing health literacy levels. The incorporation of AI LLMs into clinical practice has the potential to transform how patients learn about their eye disease and how physicians both educate and treat patients. The successful implementation of AI LLMs may lead to more personalized healthcare and informed decision-making for patients in addition to improved efficiency for providers.

## Clinical Significance

Patients are increasingly turning to the internet to learn more about their ocular disease. We showed that ChatGPT responses to glaucoma-related patient questions were more accurate than Google responses but were written at a significantly higher reading level. ChatGPT did show an ability to generate patient education materials at specified reading levels when requested, which can help physicians tailor education efforts to patients with varying health literacy. Our study demonstrates the utility of ChatGPT for patients seeking information about glaucoma and for physicians when creating unique patient education materials at reading levels that optimize understanding by patients.

## REFERENCES

1. Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. *J Med Int Res* 2017;19(1):e5729. DOI: 10.2196/jmir.5729
2. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, et al. Online health information seeking among US adults: measuring progress toward a healthy people 2020 objective. *Public Health Rep* 2019;134(6):617–625. DOI: 10.1177/0033354919874074
3. Drees J. Google receives more than 1 billion health questions every day. 2019. <https://www.beckershospitalreview.com/healthcareinformation-technology/google-receives-more-than-1-billionhealth-questions-every-day.html>. Accessed 20 Feb 2024.
4. Cohen SA, Fisher AC, Pershing S. Analysis of the readability and accountability of online patient education materials related to glaucoma diagnosis and treatment. *OPHTH* 2023;17:779–788. DOI: 10.2147/OPHTH.S401492
5. Cohen SA, Pershing S. Readability and accountability of online patient education materials for common retinal diseases. *Ophthalmol Retina* 2022;6(7):641–643. DOI: 10.1016/j.oret.2022.03.015
6. Cohen SA, Tijerina JD, Kessler A. The readability and accountability of online patient education materials related to common oculo-plastics diagnoses and treatments. *Semin Ophthalmol* 2023;38(4):387–393. DOI: 10.1080/08820538.2022.2158039
7. Hua HU, Rayess N, Li AS, et al. Quality, readability, and accessibility of online content from a google search of “macular degeneration”: critical analysis. *J Vitreoretin Dis* 2022;6(6):437–442. DOI: 10.1177/24741264221094683
8. Martin CA, Khan S, Lee R, et al. Readability and suitability of online patient education materials for glaucoma. *Ophthalmol Glaucoma* 2022;5(5):525–530. DOI: 10.1016/j.ogla.2022.03.004
9. Cohen SA, Brant A, Rayess N, et al. Google Trends—assisted analysis of the readability, accountability, and accessibility of online patient education materials for the treatment of AMD After FDA approval of pegcetacoplan. *J Vitreoretin Dis* 2024;8(4):421–427. DOI: 10.1177/24741264241250156
10. Gupta B, Mufti T, Sohail SS, et al. ChatGPT: a brief narrative review. *Cogent Bus Manag* 2023;10(3):2275851. DOI: 10.1080/23311975.2023.2275851
11. De Angelis L, Baglivo F, Arzilli G, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health* 2023;11:1166120. DOI: 10.3389/fpubh.2023.1166120
12. Rosenberg GS, Magnéli M, Barle N, et al. ChatGPT-4 generates orthopedic discharge documents faster than humans maintaining comparable quality: a pilot study of 6 cases. *Acta Orthop* 2024;95:152–156. DOI: 10.2340/17453674.2024.40182

13. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. *Semin Ophthalmol* 2023;38(5):503–507. DOI: 10.1080/08820538.2023.2209166
14. Cohen SA, Brant A, Fisher AC, et al. 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(6):472–479. DOI: 10.1080/08820538.2024.2326058
15. Bernstein IA, Zhang Y, Govil D, et al. Comparison of ophthalmologist and large language model chatbot responses to online patient eye care questions. *JAMA Network Open* 2023;6(8):e2330320. DOI: 10.1001/jamanetworkopen.2023.30320
16. Cappellani F, Card KR, Shields CL, et al. Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients. *Eye* 2024;38:1–6. DOI: 10.1038/s41433-023-02906-0
17. Potapenko I, Boberg-Ans LC, Stormly Hansen M, et al. Artificial intelligence-based chatbot patient information on common retinal diseases using ChatGPT. *Acta Ophthalmol* 2023;101(7):829–831. DOI: 10.1111/aos.15661
18. Momenaei B, Wakabayashi T, Shahlaee A, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina* 2023;7(10):862–868. DOI: 10.1016/j.oret.2023.05.022
19. Cox A, Seth I, Xie Y, et al. Utilizing ChatGPT-4 for providing medical information on blepharoplasties to patients. *Aesthet Surg J* 2023;43(8):NP658–NP662. DOI: 10.1093/asj/sjad096
20. Samaan JS, Yeo YH, Rajeev N, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. *Obes Surg* 2023;33(6):1790–1796. DOI: 10.1007/s11695-023-06603-5
21. Onder CE, Koc G, Gokbulut P, et al. Evaluation of the reliability and readability of ChatGPT-4 responses regarding hypothyroidism during pregnancy. *Sci Rep* 2024;14(1):243. DOI: 10.1038/s41598-023-50884-w
22. Sarraju A, Bruemmer D, Van Iterson E, et al. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA* 2023;329(10):842–844. DOI: 10.1001/jama.2023.1044
23. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. *Cyberpsychol Behav* 2004;7(5):497–510. DOI: 10.1089/cpb.2004.7.497
24. Wang Y, McKee M, Torbica A, et al. Systematic literature review on the spread of health-related misinformation on social media. *Soc Sci Med* 2019;240:112552. DOI: 10.1016/j.socscimed.2019.112552
25. Solomon SD, Shoge RY, Ervin AM, et al. Improving access to eye care: a systematic review of the literature. *Ophthalmology* 2022;129(10):e114–e126. DOI: 10.1016/j.ophtha.2022.07.012
26. Musa I, Bansal S, Kaleem MA. Barriers to care in the treatment of glaucoma: socioeconomic elements that impact the diagnosis, treatment, and outcomes in glaucoma patients. *Curr Ophthalmol Rep* 2022;10(3):85–90. DOI: 10.1007/s40135-022-00292-6
27. Schaeffer K. 10 facts about today's college graduates. Pew Research Center. <https://www.pewresearch.org/short-reads/2022/04/12/10-facts-about-todays-college-graduates/>. Accessed 20 Feb 2024.
28. Kianian R, Sun D, Crowell EL, et al. The use of large language models to generate education materials about uveitis. *Oph Retina* 2024;8(2):195–201. DOI: 10.1016/j.oret.2023.09.008
29. Killeen OJ, Niziol LM, Cho J, et al. Glaucoma medication adherence 1 year after the support, educate, empower personalized glaucoma coaching program. *Ophthalmol Glaucoma* 2023;6(1):23–28. DOI: 10.1016/j.ogla.2022.08.001
30. Newman-Casey PA, Niziol LM, Lee PP, et al. The impact of the support, educate, empower personalized glaucoma coaching pilot study on glaucoma medication adherence. *Ophthalmol Glaucoma* 2020;3(4):228–237. DOI: 10.1016/j.ogla.2020.04.013
31. Muir KW, Lee PP. Health literacy and ophthalmic patient education. *Surv Ophthalmol* 2010;55(5):454–459. DOI: 10.1016/j.survophthal.2010.03.005
32. Allison K, Patel DG, Greene L. Racial and ethnic disparities in primary open-angle glaucoma clinical trials: a systematic review and meta-analysis. *JAMA Network Open* 2021;4(5):e218348. DOI: 10.1001/jamanetworkopen.2021.8348
33. Hickey KT, Creber RMM, Reading M, et al. Low health literacy. *Nurse Pract* 2018;43(8):49–55. DOI: 10.1097/01.NPR.0000541468.54290.49
34. Chaudhry SI, Herrin J, Phillips C, et al. Racial disparities in health literacy and access to care among patients with heart failure. *J Card Fail* 2011;17(2):122–127. DOI: 10.1016/j.cardfail.2010.09.016
35. Redd TK, Read-Brown S, Choi D, et al. Electronic health record impact on pediatric ophthalmologists' productivity and efficiency at an academic center. *J AAPOS* 2014;18(6):584–589. DOI: 10.1016/j.jaapos.2014.08.002
36. Chiang MF, Read-Brown S, Tu DC, et al. Evaluation of electronic health record implementation in ophthalmology at an academic medical center (an American Ophthalmological Society thesis). *Trans Am Ophthalmol Soc* 2013;111:70–92. PMID: 24167326.
37. Chen AJ, Baxter SL, Gali HE, et al. Evaluation of electronic health record implementation in an academic oculoplastics practice. *Ophthalmic Plast Reconstr Surg* 2020;36(3):277–283. DOI: 10.1097/IOP.0000000000001531
38. Li Z, Wang L, Wu X, et al. Artificial intelligence in ophthalmology: the path to the real-world clinic. *Cell Rep Med* 2023;4(7):101095. DOI: 10.1016/j.xcrm.2023.101095
39. Berkowitz ST, Finn AP, Parikh R, et al. Ophthalmology workforce projections in the United States, 2020 to 2035. *Ophthalmology* 2024;131(2):133–139. DOI: 10.1016/j.ophtha.2023.09.018
40. John AM, John ES, Hansberry DR, et al. Analysis of the readability of patient education materials in pediatric ophthalmology. *JAAPOS* 2015;19(4):e48. DOI: 10.1016/j.jaapos.2015.07.149
41. Pakhchanian H, Yuan M, Raiker R, et al. Readability analysis of the American Society of ophthalmic plastic & reconstructive surgery patient educational brochures. *Semin Ophthalmol* 2022;37(1):77–82. DOI: 10.1080/08820538.2021.1919721