

# Comparing the Ability of Google and ChatGPT to Accurately Respond to Oculoplastics-Related Patient Questions and Generate Customized Oculoplastics Patient Education Materials

Samuel A Cohen<sup>1</sup>, Nikhita Yadlapalli<sup>2</sup>, Jonathan D Tijerina<sup>3</sup>, Chrisfouad R Alabiad<sup>3</sup>, Jessica R Chang<sup>4</sup>, Benyam Kinde<sup>5</sup>, Nicholas R Mahoney<sup>6</sup>, Kelsey A Roelofs<sup>1</sup>, Julie A Woodward<sup>7</sup>, Andrea L Kossler<sup>5</sup> 

<sup>1</sup>Department of Ophthalmology, Stein Eye Institute at University of California Los Angeles David Geffen School of Medicine, Los Angeles, CA, USA; <sup>2</sup>Department of Ophthalmology, FIU Herbert Wertheim College of Medicine, Miami, FL, USA; <sup>3</sup>Department of Ophthalmology, Bascom Palmer Eye Institute at University of Miami Miller School of Medicine, Miami, FL, USA; <sup>4</sup>Department of Ophthalmology, USC Roski Eye Institute at University of Southern California Keck School of Medicine, Los Angeles, CA, USA; <sup>5</sup>Department of Ophthalmology, Byers Eye Institute at Stanford University School of Medicine, Palo Alto, CA, USA; <sup>6</sup>Department of Ophthalmology, Wilmer Eye Institute at Johns Hopkins University School of Medicine, Baltimore, MD, USA; <sup>7</sup>Department of Ophthalmology, Duke Eye Center at Duke University School of Medicine, Durham, NC, USA

Correspondence: Andrea L Kossler, Department of Ophthalmology, Byers Eye Institute at Stanford University School of Medicine, 2452 Watson Court, Palo Alto, CA, USA, Email [akossler@stanford.edu](mailto:akossler@stanford.edu)

**Purpose:** To compare the accuracy and readability of responses to oculoplastics patient questions provided by Google and ChatGPT. Additionally, to assess the ability of ChatGPT to create customized patient education materials.

**Methods:** We executed a Google search to identify the 3 most frequently asked patient questions (FAQs) related to 10 oculoplastics conditions. FAQs were entered into both the Google search engine and the ChatGPT tool and responses were recorded. Responses were graded for readability using five validated readability indices and for accuracy by six oculoplastics surgeons. ChatGPT was instructed to create patient education materials at various reading levels for 8 oculoplastics procedures. The accuracy and readability of ChatGPT-generated procedural explanations were assessed.

**Results:** ChatGPT responses to patient FAQs were written at a significantly higher average grade level than Google responses (grade 15.6 vs 10.0,  $p < 0.001$ ). ChatGPT responses (93% accuracy) were significantly more accurate ( $p < 0.001$ ) than Google responses (78% accuracy) and were preferred by expert panelists (79%). ChatGPT accurately explained oculoplastics procedures at an above average reading level. When instructed to rewrite patient education materials at a lower reading level, grade level was reduced by approximately 4 (15.7 vs 11.7, respectively,  $p < 0.001$ ) without sacrificing accuracy.

**Conclusion:** ChatGPT has the potential to provide patients with accurate information regarding their oculoplastics conditions. ChatGPT may also be utilized by oculoplastic surgeons as an accurate tool to provide customizable patient education for patients with varying health literacy. A better understanding of oculoplastics conditions and procedures amongst patients can lead to informed eye care decisions.

**Keywords:** oculoplastics, google, ChatGPT, readability, accuracy

## Introduction

The internet remains a vital data source for patients seeking health information. Recent estimates suggest that approximately 80% of Americans utilize the internet for health searches.<sup>1</sup> Google is the most common search engine used by patients when seeking health information – as nearly 7% of Google's more than 8.5 billion daily searches are health-related.<sup>2</sup> Given Google's role as the predominant search engine used by patients, it is imperative that responses to patient questions are accurate and understandable. Yet, prior research has demonstrated that Google search information related to

ophthalmologic subspecialties including glaucoma, retina, and oculoplastics is often inaccurate and written at a reading level that may be difficult to understand by patients.<sup>3–7</sup>

In the past year, artificial intelligence (AI) large language models (LLMs) have risen in popularity amongst the general public and have emerged as an alternative information source for patients.<sup>8</sup> ChatGPT has become the most popular LLM in the US, amassing more than 200 million users since 2022.<sup>9,10</sup> Prior studies have shown that the ability of ChatGPT to respond to patient queries has been mixed in ophthalmology and in medicine as a whole – with some studies reporting that responses provided by ChatGPT were accurate while others reported that ChatGPT responses contained inaccurate and potentially harmful information.<sup>11–13</sup>

In addition to serving as an information source for patients, LLMs like ChatGPT also have the potential to greatly enhance patient education capabilities for physicians. LLMs which demonstrate an ability to easily break down complex medical phenomenon into terms that patients can better understand may contribute to more informed shared decision making between physicians and patients, which can ultimately improve patient outcomes.<sup>14–17</sup>

While the ability of both Google and LLMs like ChatGPT to respond to patient queries has been explored in several ophthalmology subspecialties, to date, there is limited information on this topic in oculoplastics.<sup>17–20</sup> Furthermore, it is unclear whether ChatGPT has the ability to adjust patient education materials to appropriate reading levels while maintaining accuracy of information.

As such, the purpose of our study is twofold. First, a panel of fellowship-trained oculoplastics surgeons will compare the accuracy and readability of responses to common oculoplastics patient questions provided by two information sources: Google and ChatGPT. Second, we assess the ability of ChatGPT to assist physicians in creating customized patient education materials to better explain common oculoplastics procedures to patients with varying health literacy.

## Materials and Methods

### Generation and Responses of FAQs

We used a clean-installed Google Chrome (Menlo Park, CA) browser on Incognito Mode to execute all Google searches used in data collection. We disabled location filters and sponsored results to avoid bias from previous searches and targeted geographic search results. The 10 search terms used in this study were identified from a prior study evaluating patient education materials for oculoplastics diagnoses, which was based on the American Society of Ophthalmic Plastic and Reconstructive Surgery (ASOPRS) website.<sup>4</sup> The search terms included: “thyroid eye disease”, “orbital cellulitis”, “orbital tumor”, “proptosis”, “ptosis”, “entropion”, “blepharospasm”, “chalazion”, “nasolacrimal duct obstruction”, and “epiphora”.

After each of the 10 Google searches were executed, the first three “Frequently Asked Questions” (FAQs) associated with each of the search terms were recorded. These 30 FAQs were entered into both the Google search engine and the ChatGPT V 3.5 tool and responses were recorded. For FAQs entered into Google, the response that was recorded was the response processed by the Google search engine that populated on the Google search page itself. However, the Google response which populates was derived from a website, and the website used to generate the response was also noted. For all ChatGPT queries in our study, the chatbot was instructed to respond with a temperature setting of 0.3 (scale 0–1, with 0 more focused responses and 1 more creative responses) to ensure more focused responses intended for use by physicians and patients.

### Readability Analysis

Responses to patient FAQs provided by both Google and ChatGPT were evaluated for readability using 5 validated, objective 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). The FRE scale measures readability by generating a score from 0 (very difficult) to 100 (very easy). Each of the four remaining readability scales provide a “grade-level” at which the article was written. For example, a score of 6 indicates the text was written at a 6th grade reading level. An average grade level for each article was calculated from the GFI, FKGL, SMOG, and CLI indices.

## JAMA Accountability Analysis

All 30 websites used by Google to generate responses to patient FAQs were evaluated for accountability (score of 0–4) using *JAMA* benchmarks. According to *JAMA* guidelines, a website containing patient education materials should (1) include all authors and their relevant credentials (2) list references (3) provide disclosures and (4) provide date of last update.

## ChatGPT Procedure Explanations

ChatGPT was instructed to generate patient education materials for 8 common oculoplastics procedures to help patients better understand each respective procedure. The 8 procedures included: blepharoplasty, browplasty, dacryocystorhinostomy, rhinoplasty, orbital decompression, levator advancement, ectropion repair, and botox for blepharospasm. 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 the five aforementioned 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 patient education materials at a 7<sup>th</sup> grade reading level. The phrase, “ChatGPT, please explain what happens in a [procedure name] at a 7<sup>th</sup> grade reading level” was entered into the ChatGPT tool. The responses provided were recorded and, again, evaluated for readability.

## Expert Panel Evaluation - FAQs

A panel of 6 fellowship-trained oculoplastics surgeons employed by six different academic institutions were asked to independently review responses to each of the 30 FAQs generated by both Google and AI for several different criteria. For all 30 FAQs, the responses provided by Google and AI were listed side-by-side in a randomized order and three questions were asked. Experts were first asked to select the better (more accurate and comprehensible) response to the patient question. They were then asked to identify which response was generated by artificial intelligence. Finally, they were asked if either of the responses contained inappropriate or inaccurate information.

## Expert Panel Evaluation – Procedures

For each of the 8 oculoplastics procedures, the explanations provided by ChatGPT at various reading levels (unspecified and 7<sup>th</sup> grade level request) were listed side-by-side in a randomized order. Experts were 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 the block of text written at a lower reading level. Finally, they were asked if either of the responses contained inappropriate or inaccurate information.

## Statistical Analysis

Data was analyzed in R version 4.3.2. Two-sample *t*-tests were utilized to compare 1) the readability of responses provided by Google and ChatGPT and 2) the readability of procedure explanations provided by ChatGPT at various requested reading levels. 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.

## Results

### Frequently Asked Questions Related to Common Oculoplastics Conditions

Table 1 shows the 30 FAQs that populated after executing a Google search for 10 common oculoplastics conditions. Both open-ended and closed-ended questions populated, with qualitative and quantitative patient inquiries.

### Readability of Responses to Patient FAQs - Google vs ChatGPT

Responses to patient FAQs provided by ChatGPT were written at a significantly higher average grade level than responses provided by Google (15.6 vs 10.0, respectively,  $p < 0.001$ ) (Table 2). ChatGPT responses also averaged a lower (more difficult to understand) FRE than Google responses (27.4 vs 48.5, respectively,  $p < 0.001$ ).

**Table 1** Most Frequently Asked Questions for 10 Common Oculoplastics Conditions

Category	Frequently Asked Question 1	Frequently Asked Question 2	Frequently Asked Question 3
Thyroid Eye Disease	What is thyroid eye disease?	What are the first signs of thyroid eye disease?	What triggers thyroid eye disease?
Orbital Cellulitis	What is the main cause of orbital cellulitis?	What are the first signs of orbital cellulitis?	How do you get orbital cellulitis?
Orbital Tumors	What is the most common orbital tumor in adult patients?	What is the prognosis for orbital tumors?	How do you get rid of orbital tumors?
Proptosis	What is the most common cause of proptosis in adults?	What is proptosis?	Can you fix proptosis?
Ptosis	What is ptosis?	What happens if ptosis is left untreated?	How do you fix ptosis without surgery?
Entropion	What is entropion?	How is entropion treated?	What is the recovery time for entropion?
Blepharospasm	What is blepharospasm?	How do you fix blepharospasm?	What is the main cause of blepharospasm?
Chalazion	What is a chalazion?	How do you get rid of a chalazion?	What is a chalazion filled with?
Nasolacrimal Duct Obstruction	What is nasolacrimal duct obstruction?	How do you treat nasolacrimal duct blockage?	What are the symptoms of a blocked tear duct?
Epiphora	What is epiphora?	How do you treat epiphora in the eye?	How do you test for epiphora?

**Table 2** Readability of Responses to Oculoplastics FAQs – Google Vs ChatGPT

	Flesch-Kincaid 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	48.5	8.4	8.9	7.2	15.6	10.0
ChatGPT	27.4	14.2	18.1	12.9	17.2	15.6

Notes: \*Differences in both FRE and Average Grade Level statistically significant at  $p < 0.001$ .

## JAMA Accountability of Webpages Providing Answers to FAQs

Most webpages providing answers to Google FAQs related to the oculoplastics conditions were from educational institutions, with fewer webpages from national organizations and private practices. The average *JAMA* accountability score of the 30 webpages analyzed was 1.07/4. The most frequent *JAMA* accountability metrics on webpages analyzed included date of last update (50%), author list (20%), and references (17%).

## Readability of ChatGPT Generated Patient Education Materials for Common Oculoplastics Procedures

Table 3 displays the readability metrics of the ChatGPT-generated patient education materials for common oculoplastics procedures. Without specifying a grade level at which patient education materials should be written, the average grade

**Table 3** Readability of Oculoplastics Procedure Explanations – Unspecified Grade Level vs 7th Grade Level Requested

	Flesch-Kincaid 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	29.8	14.5	18.6	13.5	16.3	15.7
7 <sup>th</sup> Grade Level	53.6	10.4	13.4	9.9	13.0	11.7

Notes: \*Differences in both FRE and Average Grade Level statistically significant at  $p < 0.001$ .

level of the patient education materials produced was 15.7 (college level), with a FRE of 29.8 (“very difficult”). When asking ChatGPT to generate patient education materials at a 7<sup>th</sup> grade reading level, readability significantly improved to an average grade level of 11.7 (high school level) and a FRE of 53.6 (“fairly difficult”) ( $p < 0.001$ ).

## Expert Panel Evaluation – Google and ChatGPT Responses to FAQs

When asked to identify the better (more accurate and comprehensible) response to the 30 FAQs included in our study, responses provided by ChatGPT were selected 79% (142/180) of the time by expert reviewers while Google responses were selected less frequently (12%, 21/180). Table 4 shows the distribution of how expert reviewers compared AI and Google responses to patient FAQs by topic.

When asked to identify which of the provided responses to the 30 FAQs was generated by AI, panelists were largely unsuccessful. Approximately 33% of responses to this question (59/180) were, “I cannot tell which response was generated by an artificial intelligence large language model”. Of the panelists who elected to make a selection, 31% (38/121) correctly identified the AI-generated response (Table 5).

When asked whether the AI-generated or Google-generated responses to patient FAQs contained inaccurate or inappropriate information, the majority of responses did not contain inaccurate or inappropriate information (136/180, 76%). Responses provided by Google were rated to contain inaccurate or inappropriate information 22% of the time while responses provided by ChatGPT were rated to contain inaccurate or inappropriate information 7% of the time ( $p < 0.001$ ).

## Expert Panel Evaluation – ChatGPT Generated Patient Education Materials for Common Oculoplastics Procedures

When asked to select which block of text (unspecified grade level vs 7<sup>th</sup> grade reading level requested) they would use to incorporate into patient education materials about common oculoplastics procedures, responses were distributed between the text written at the unspecified grade level (19/48, 40%), 7<sup>th</sup> grade reading level (12/48, 25%), and both equally (17/48, 35%) (Table 6). Panelists were accurately able to select which of the responses was written at the lower reading level for most procedures (30/48, 63%). When asked to evaluate whether each block of text contained inaccurate or inappropriate information, “neither” was selected 94% of the time. Both the responses provided at the unspecified grade level and the responses generated after the 7<sup>th</sup> grade reading level was requested were labeled as inaccurate on 3% of total panelist reviews.

**Table 4** Expert Review of the Better (More Accurate and Comprehensible) Responses to Patients’ Frequently Asked Questions by Topic (3 Questions for Each Topic Reviewed by All 6 Panelists)

Topic	AI	Google	Both Answers Equally Appropriate
Thyroid Eye Disease	15	1	2
Orbital Cellulitis	18	0	0
Orbital Tumors	16	1	1
Proptosis	14	1	3
Ptosis	11	6	1
Entropion	12	2	4
Blepharospasm	15	2	1
Chalazion	15	2	1
Nasolacrimal Duct Obstruction	15	2	1
Epiphora	11	4	3
<i>Total</i>	142	21	17

**Abbreviation:** AI, Artificial Intelligence.

**Table 5** Expert Panelist Ability to Correctly Identify Responses to Patients' Frequently Asked Questions Generated by Artificial Intelligence 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
Thyroid Eye Disease	5	9	4
Orbital Cellulitis	4	11	3
Orbital Tumors	4	7	7
Proptosis	2	10	6
Ptosis	5	8	5
Entropion	6	5	7
Blepharospasm	4	5	9
Chalazion	3	9	6
Nasolacrimal Duct Obstruction	2	10	6
Epiphora	3	9	6
<i>Total</i>	38	83	59

**Abbreviation:** AI, Artificial Intelligence.

**Table 6** Reading Level of Patient Education Information Preferred by Oculoplastics Surgeons if They Were Creating Informational Pamphlets for Patients by Topic

Topic	Unspecified Reading Level (Average Grade Level 15.7)	7 <sup>th</sup> Grade Reading Level Requested (Average Grade Level 11.7)	Both Equally
Blepharoplasty	5	0	1
Browplasty	0	3	3
DCR	4	1	1
Rhinoplasty	2	1	3
Orbital Decompression	0	3	3
Levator Advancement	1	2	3
Ectropion Repair	4	0	2
Botox	3	2	1
<i>Total</i>	19	12	17

## Discussion

Our study sought to compare the responses generated by Google and an artificial intelligence large language model (ChatGPT) to FAQs regarding common oculoplastics conditions. Additionally, our study assessed the efficacy of ChatGPT in generating patient-oriented educational materials tailored to varying levels of health literacy for common oculoplastics procedures. Ideally, patient education materials are written at a low grade level with high accuracy, thereby achieving maximum reach. We found that when responding to patient FAQs, responses generated by ChatGPT were rated as more accurate and were preferred by a panel of fellowship-trained oculoplastics surgeons. However, responses provided by ChatGPT were written at a significantly higher grade level than responses generated by Google, which could impair understandability among patients with lower health literacy. ChatGPT was able to effectively create accurate patient education materials containing information about several common oculoplastics procedures. Furthermore, if prompted to generate patient education materials at a lower reading level, ChatGPT was able to effectively reduce the grade level at which procedure explanations were written from a college grade level to a high school grade level without sacrificing accuracy. Large language models such as ChatGPT have the potential to revolutionize both the way patients learn about their health and the way physicians educate patients. Oculoplastic surgeons cognizant of the many potential applications of LLMs, such as ChatGPT, may utilize these tools to ensure patients receive accurate information online while also tailoring education interventions to patients with varying health literacy.

Our study adds to the existing body of literature regarding the use of LLMs in ophthalmology. While a recent study determined that ChatGPT produces incomplete and potentially harmful information about common ophthalmic conditions,<sup>13</sup> our results support two prior studies which found that ChatGPT was largely effective in responding to patients' eye care questions.<sup>18,21</sup> The authors believe that LLMs represent an accurate information source for patients and can be utilized by providers as an effective patient educational tool.

With patients increasingly turning to the internet to learn more about their health, our study's findings have important implications for the future of patient education. Currently, Google is the most common website patients turn to for health information.<sup>22</sup> Our results demonstrate that 6 expert oculoplastic surgeons considered ChatGPT responses to patient queries more accurate than Google search responses. These results are interesting as most Google Search responses came directly from a physician website. Furthermore, oculoplastics surgeons had difficulty distinguishing AI-generated from Google-generated responses, with less than 50% accuracy overall, likely indicating preconceived notions about the types of responses likely to be generated by LLMs vs Google. LLMs have the potential to simplify and improve the way oculoplastic surgeons and their teams provide patient information. Physicians can use LLMs to update patient facing educational materials in clinic and on their websites by making them more understandable for patients with varying health literacy. Additionally, electronic health record systems can integrate AI tools as a first response option to common patient questions in an effort to minimize physician extender, registered nurse, and physician phone calls and charting time. Further, patients who utilize LLMs for health information also benefit from receiving a direct response to their inquiry, rather than having to sift through the vast amount of misinformation that patients often encounter after executing a Google search with a medical question.<sup>23</sup>

While ChatGPT may represent an accurate alternative for patient inquiries regarding common oculoplastics procedures, readability of responses by ChatGPT were far higher than the 6<sup>th</sup> grade level recommended by the American Medical Association. Average responses to patient queries in our study were written at a 16<sup>th</sup> grade level (college level), which may be difficult to understand for patients without a college degree (62% of Americans).<sup>24</sup> Oculoplastics surgeons must therefore exercise caution when referring patients to ChatGPT as a potential information source. Patients should be directed to specifically note that they prefer a lower reading level when executing their ChatGPT inquiry in order to increase the chance that the response provided by the tool is at an appropriate reading level. There are, however, limitations with this approach. When instructed to provide responses at a 7<sup>th</sup> grade reading level, the responses provided by ChatGPT were written at a 12<sup>th</sup> grade level – an improvement compared to when no grade level was specified but still higher than the 7<sup>th</sup> grade reading level requested. Further advances in LLMs are required to more accurately provide information at a requested reading level.

When using LLM for the purpose of creating patient facing educational materials, oculoplastics surgeons should understand ChatGPT's ability to tailor patient education materials to lower reading levels without sacrificing accuracy – and should therefore specify their target reading level. In our study, the current version of ChatGPT was effective in improving understandability of text from a college graduate level to a high school graduate level, which would make patient education materials far more inclusive, as approximately 87% of Americans have obtained a high school degree as opposed to only 38% with a college degree.<sup>24,25</sup> Furthermore, survey respondents indicated that they were comfortable sharing patient education materials written at a lower reading level with patients. This is important because prior research in ophthalmology shows that tailoring patient educational materials to a patient's health literacy level can empower patients to make informed eye care decisions and ultimately result in improved patient outcomes.<sup>26–28</sup> Additionally, the use of LLMs to translate patient education materials to other languages could help to address language barriers in healthcare that often result in worse patient outcomes.<sup>29,30</sup> Oculoplastics surgeons who are familiar with the ability of ChatGPT to customize patient education materials for patients with varying health literacy levels can use the tool to help ensure that all patients understand the basics of their eye condition when engaging in shared decision making regarding treatment options.

There are limitations to our study. First, the 10 oculoplastics conditions and 8 oculoplastics procedures utilized in this study represent a small fraction of all oculoplastics conditions and procedures, respectively. As such, our findings may not be generalizable to patient education regarding other conditions and procedures not included in this study. However, both the conditions and procedures selected for inclusion in this study were chosen based on prior research, public interest based on Google search volume, and author consensus regarding the most frequent conditions and procedures, respectively, they encounter in everyday practice.<sup>4,31</sup> Next, while expert reviewers in our study are employed by six

different academic institutions, these reviewers represent a small sample of all oculoplastics surgeons in the country. As such, subjective survey responses may not be representative of the opinions of all oculoplastics surgeons, given the wide variability in responses provided by expert reviewers. Next, details regarding the data sets used to train the ChatGPT LLM as well as its intended target audience are not publicly available. Therefore, it is plausible that some patient education websites that populate after Google search could have been used to “teach” ChatGPT how to respond to patient queries. The impact that Google search results have on learned ChatGPT responses is unknown at this time. Furthermore, our study utilized ChatGPT version 3.5 – it would be beneficial to further investigate future versions of ChatGPT to determine if improvements to the interface can continue to optimize patient education. Additionally, we utilized the Google Chrome browser to execute the Google search to provide answers to the 30 FAQs generated for this study. It is possible that other browsers may have populated different answers to the 30 FAQs; however, a repeat analysis of the answers which populated after Google search on the Microsoft Edge browser showed similar website links. Finally, while we used readability indices as a proxy for understandability of answers to patient queries and procedural explanations, the terms “readability” and “understandability” are not synonymous. It is possible that a patient may consider a response that is more “readable” to be more difficult to understand when compared with another response that is less “readable”; however, the validated readability indices used in this study are frequently utilized as a barometer of understandability of patient education materials for patients in both ophthalmology and other medical fields.<sup>32–35</sup>

## Conclusions

In conclusion, expert reviewers in our study found ChatGPT responses to oculoplastics FAQs to be more accurate when compared with Google responses, suggesting that ChatGPT has the potential to become a useful resource for patients seeking information about their oculoplastics conditions. Furthermore, we demonstrate the ability for ChatGPT to improve patient education materials by reducing the grade level at which procedural explanations are written, which can help oculoplastics surgeons tailor patient education efforts to patients with varying health literacy. A better understanding of oculoplastics conditions and procedures amongst patients can help lead to informed eye care decisions and may enhance patient satisfaction and outcomes.

## Funding

This work was supported by an unrestricted grant from research to prevent blindness and the National Institute of Health grant, NIH P30 026877.

## Disclosure

The authors declare no conflicts of interest.

## References

1. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, Allen SV, Moser RP, Hesse BW. 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
2. Drees J Google receives more than 1 billion health questions every day. 2019. Available from: <https://www.beckershospitalreview.com/healthcare-information-technology/google-receives-more-than-1-billion-health-questions-every-day.html>. Accessed February 20, 2024.
3. Cohen SA, Fisher AC, Pershing S. Analysis of the readability and accountability of online patient education materials related to glaucoma diagnosis and treatment. *Clin Ophthalmol.* 2023;17:779–788. doi:10.2147/OPTH.S401492
4. Cohen SA, Tijerina JD, Kossler A. The readability and accountability of online patient education materials related to common oculoplastics diagnoses and treatments. *Semin Ophthalmol.* 2023;38(4):387–393. doi:10.1080/08820538.2022.2158039
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. Williams AM, Muir KW, Rosdahl JA. Readability of patient education materials in ophthalmology: a single-institution study and systematic review. *BMC Ophthalmol.* 2016;16(1):133. doi:10.1186/s12886-016-0315-0
7. Crabtree L, Lee E. Assessment of the readability and quality of online patient education materials for the medical treatment of open-angle glaucoma. *BMJ Open Ophthalmol.* 2022;7(1):e000966. doi:10.1136/bmjophth-2021-000966
8. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. *Front Artif Intell.* 2023;6:1169595. doi:10.3389/frai.2023.1169595
9. Peres R, Schreier M, Schweidel D, Sorescu A. On ChatGPT and beyond: how generative artificial intelligence may affect research, teaching, and practice. *Int J Mark Res.* 2023;40(2):269–275. doi:10.1016/j.ijresmar.2023.03.001
10. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *J Med Internet Res.* 2023;25:e47184. doi:10.2196/47184



11. Xie Y, Seth I, Hunter-Smith DJ, Rozen WM, Ross R, Lee M. Aesthetic surgery advice and counseling from artificial intelligence: a rhinoplasty consultation with ChatGPT. *Aesth Plast Surg*. 2023;47(5):1985–1993. doi:10.1007/s00266-023-03338-7
12. 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
13. Cappellani F, Card KR, Shields CL, Pulido JS, Haller JA. Reliability and accuracy of artificial intelligence ChatGPT in providing information on ophthalmic diseases and management to patients. *Eye*. 2024:1–6. doi:10.1038/s41433-023-02906-0
14. Waddell A, Lennox A, Spassova G, Bragge P. Barriers and facilitators to shared decision-making in hospitals from policy to practice: a systematic review. *Implement Sci*. 2021;16(1):74. doi:10.1186/s13012-021-01142-y
15. Davidson KW, Mangione CM, Barry MJ, US Preventive Services Task Force. Collaboration and shared decision-making between patients and clinicians in preventive health care decisions and us preventive services task force recommendations. *JAMA*. 2022;327(12):1171–1176. doi:10.1001/jama.2022.3267
16. Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. *Pak J Med Sci*. 2023;39(2):605–607. doi:10.12669/pjms.39.2.7653
17. Potapenko I, Boberg-Ans LC, Stormly Hansen M, Klefter ON, van Dijk EHC, Subhi Y. 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. 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
19. 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
20. Cox A, Seth I, Xie Y, Hunter-Smith DJ, Rozen WM. Utilizing ChatGPT-4 for providing medical information on blepharoplasties to patients. *Aesthet Surg J*. 2023;43(8):NP658–NP662. doi:10.1093/asj/sjad096
21. Cohen SA, Brant A, Fisher AC, Pershing S, Do D, Pan CD. 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*. 2024:1–8. doi:10.1080/08820538.2024.2326058
22. Bachl M, Link E, Mangold F, Stier S. Search engine use for health-related purposes: behavioral data on online health information-seeking in Germany. *Health Communication*. 2024:1–14. doi:10.1080/10410236.2024.2309810
23. Wang Y, McKee M, Torbica A, Stuckler D. 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
24. Schaeffer K 10 facts about today's college graduates. pew research center. Available from: <https://www.pewresearch.org/short-reads/2022/04/12/10-facts-about-todays-college-graduates/>. Accessed February 20, 2024.
25. The NCES fast facts tool provides quick answers to many education questions (National Center for Education Statistics). Available from: <https://nces.ed.gov/fastfacts/display.asp?id=805>. Accessed February 20, 2024.
26. 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
27. Newman-Casey PA, Niziol LM, Lee PP, Musch DC, Resnicow K, Heisler M. 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
28. 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
29. de Moissac D, Bowen S. Impact of language barriers on quality of care and patient safety for official language minority francophones in Canada. *J Patient Exp*. 2019;6(1):24–32. doi:10.1177/2374373518769008
30. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a systematic review. *Oman Med J*. 2020;35(2):e122. doi:10.5001/omj.2020.40
31. Pakhchanian H, Yuan M, Raiker R, Waris S, Geist C. 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
32. Cheng BT, Kim AB, Tanna AP. Readability of online patient education materials for glaucoma. *J Glaucoma*. 2022;31(6):438–442. doi:10.1097/IJG.0000000000002012
33. John AM, John ES, Hansberry DR, Guo S. Analysis of the readability of patient education materials in pediatric ophthalmology. *J AAPOS*. 2015;19(4):e48. doi:10.1016/j.jaapos.2015.07.149
34. Patel AJ, Kloosterboer A, Yannuzzi NA, Venkateswaran N, Sridhar J. Evaluation of the content, quality, and readability of patient accessible online resources regarding cataracts. *Semin Ophthalmol*. 2021;36(5–6):384–391. doi:10.1080/08820538.2021.1893758
35. Storino A, Castillo-Angeles M, Watkins AA, et al. Assessing the accuracy and readability of online health information for patients with pancreatic cancer. *JAMA Surg*. 2016;151(9):831–837. doi:10.1001/jamasurg.2016.0730

## Clinical Ophthalmology

Dovepress

### Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-ophthalmology-journal>