AMERICAN ACADEMY OF OPHTHALMOLOGY®

Evaluation of the Appropriateness and Readability of ChatGPT-4 Responses to Patient Queries on Uveitis

S. Saeed Mohammadi, MD,[1] Anadi Khatri, MD,[1,2,3] Tanya Jain, MBBS, DNB,[1,4] Zheng Xian Thng, MD,[1,5] Woong-sun Yoo, MD, PhD,[1,6] Negin Yavari, MD,[1] Vahid Bazojoo, MD,[1] Azadeh Mobasserian, MD,[1] Amir Akhavanrezayat, MD,[1] Ngoc Trong Tuong Than, MD,[1] Osama Elaraby, MD,[1] Battuya Ganbold, MD,[1,7] Dalia El Feky, MD,[1,8] Ba Trung Nguyen, MD,[1,9] Cigdem Yasar, MD,[1] Ankur Gupta, MD, MS,[1] Jia-Horung Hung, MD,[1] Quan Dong Nguyen, MD, MSc[1]

Purpose: To compare the utility of ChatGPT-4 as an online uveitis patient education resource with existing patient education websites.

Design: Evaluation of technology.

Participants: Not applicable.

Methods: The term "uveitis" was entered into the Google search engine, and the first 8 nonsponsored websites were selected to be enrolled in the study. Information regarding uveitis for patients was extracted from Healthline, Mayo Clinic, WebMD, National Eye Institute, Ocular Uveitis and Immunology Foundation, American Academy of Ophthalmology, Cleveland Clinic, and National Health Service websites. ChatGPT-4 was then prompted to generate responses about uveitis in both standard and simplified formats. To generate the simplified response, the following request was added to the prompt: 'Please provide a response suitable for the average American adult, at a sixth-grade comprehension level.' Three dual fellowship-trained specialists, all masked to the sources, graded the appropriateness of the contents (extracted from the existing websites) and responses (generated responses by ChatGPT-4) in terms of personal preference, comprehensiveness, and accuracy. Additionally, 5 readability indices, including Flesch Reading Ease, Flesch–Kincaid Grade Level, Gunning Fog Index, Coleman–Liau Index, and Simple Measure of Gobbledygook index were calculated using an online calculator, Readable.com, to assess the ease of comprehension of each answer.

Main Outcome Measures: Personal preference, accuracy, comprehensiveness, and readability of contents and responses about uveitis.

Results: A total of 497 contents and responses, including 71 contents from existing websites, 213 standard responses, and 213 simplified responses from ChatGPT-4 were recorded and graded. Standard ChatGPT-4 responses were preferred and perceived to be more comprehensive by dually trained (uveitis and retina) specialist ophthalmologists while maintaining similar accuracy level compared with existing websites. Moreover, simplified ChatGPT-4 responses matched almost all existing websites in terms of personal preference, accuracy, and comprehensiveness. Notably, almost all readability indices suggested that standard ChatGPT-4 responses demand a higher educational level for comprehension, whereas simplified responses required lower level of education compared with the existing websites.

Conclusions: This study shows that ChatGPT can provide patients with an avenue to access comprehensive and accurate information about uveitis, tailored to their educational level.

Financial Disclosure(s): The author(s) have no proprietary or commercial interest in any materials discussed in this article. *Ophthalmology Science 2025;5:100594 © 2024 Published by Elsevier Inc. on behalf of the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Supplemental material available at www.ophthalmologyscience.org.

Uveitis is characterized by inflammation of the uvea, including the iris, ciliary body, and choroid, and poses a significant risk to vision. It affects over 2 million people globally and is the third leading cause of often preventable blindness worldwide.[1,2] It has an incidence of approximately 15 cases per 100 000 individuals annually and a prevalence of about 38 cases per 100 000 individuals in the United States.[3]

Health literacy is the extent to which an individual can acquire and comprehend health-related information and concepts, enabling them to make an informed decision.[4] The

https://doi.org/10.1016/j.xops.2024.100594
ISSN 2666-9145/24

role of health literacy in managing complex diseases is crucial, as it enables patients to effectively engage in self-care activities.[5,6] Patients with limited health literacy not only experience higher rates of hospitalization but also have been known to have an increased incidence of diseases and elevated mortality rates. They are more likely to use emergency departments for primary care needs, imposing substantial financial and social burden on the health care system.[7] A significant portion of uveitis patients—43%—lack sufficient health literacy.[8] The health literacy of uveitis patients does not correlate with the number of health care providers they consult for managing their condition, even though they have more interactions with the health care system.[8] Furthermore, previous research has shown that having a chronic disease, receiving regular eye care, or frequent clinic visits in the past year do not necessarily lead to a better understanding of the illness among patients.[9,10] On the contrary, enhanced levels of health literacy lead to improved treatment compliance, more effective communication between patients and health care providers, better mental health, and increased patient satisfaction.[11,12]

A major challenge in patient education is the literacy level of available materials. Many online resources are written at a high school or college level, which is difficult for the average US resident, who typically reads at or below an eighth-grade level.[13] This issue is more pronounced for patients covered by Medicare and Medicaid, who often read at levels below the sixth-grade.[13,14]

ChatGPT, developed by OpenAI, uses advanced deep learning techniques to generate text that closely resembles natural human language.[15,16] Its ability to generate cohesive and grammatically correct text marks a significant advancement in artificial intelligence. ChatGPT has been evaluated for its accuracy in the diagnosis and management of uveitis[17] and has been used to improve the readability of online health information about uveitis.[14] Additionally, it has been instrumental in question-answering, delivering trustworthy information about various diseases and medical inquiries.[18,19] The effectiveness of ChatGPT in generating easier-to-understand uveitis health information for patients, compared with existing websites, has been assessed.[14] However, the appropriateness and readability of its responses to patient inquiries about uveitis remain unevaluated. Such a comparison is essential in evaluating the utility of ChatGPT in enhancing patient understanding and engagement in their health care, especially considering the prevalent issue of low health literacy.

In this study, we aim to compare the appropriateness and readability of ChatGPT-4 responses to patient queries about uveitis with respect to various existing websites.

Methods

Data Collection

In November 2023, the search term "uveitis" was entered into the Google search engine, and the first 8 nonsponsored websites appearing on the first search page were selected for enrollment in the study. These websites included Healthline, Mayo Clinic, WebMD, National Eye Institute (NEI), Ocular Uveitis and Immunology Foundation, American Academy of Ophthalmology (AAO), Cleveland Clinic, and the National Health Service. All these sources have dedicated webpages for patient education regarding uveitis, where they explain the disease in simpler terms for patients. On each webpage, there are several titles with corresponding explanation (e.g., under the title "What is uveitis?" there is an explanation about it). Each title and its corresponding explanation were extracted and referred to as "content" in this manuscript. Links to each webpage and the title of each content are shown in Table S1.

Standard ChatGPT-4 Responses

If the title of a content was in a question format, the same question was then queried from ChatGPT-4 3 times, and responses were recorded. If the title was not initially in question format, it was reformulated accordingly (e.g., if a header was “symptoms of uveitis” on the website, it was reformatted to ’What are the symptoms of uveitis?’).

Simplified ChatGPT-4 Responses

To simulate simplified ChatGPT-4 responses, the same queries were entered into ChatGPT-4 for 3 times with an additional prompt: ‘Please provide a response suitable for the average American adult, at a sixth-grade comprehension level.’

Uniforming the Data

The extracted contents and responses were then made similar in terms of font and text size to ensure consistency. Subsequently, the data were deidentified by removing any identifiers pointing to websites, institutes, or ChatGPT-4 throughout the text, and a unique code was assigned to each specific content for analysis purposes.

Evaluation of Appropriateness by Uveitis and Retina Specialists

Three dual fellowship-trained specialists (A.K. and T.J., with fellowship training in vitreoretinal surgery and uveitis, and Z.X.T., with fellowship training in medical retina and uveitis), who were masked to the source of the texts, evaluated the appropriateness of the contents and responses. They were tasked with grading the texts using a 5-point Likert scale, where 1 represented "strongly disagree" and 5 denoted "strongly agree." Appropriateness was evaluated in terms of personal preference, comprehensiveness, and accuracy of the contents and responses. This scale allowed for a nuanced assessment of the quality and relevance from a specialist’s perspective. Grades assigned by these experts were then averaged and analyzed.

Readability Assessment

The readability of each piece of uniformed content or response was assessed using Readable.com, an online readability calculator.[20] This process involved analyzing the texts to count the number of characters, words, syllables, and sentences in order to assess the ease of comprehension and level of education needed to understand the text. Subsequently, 5 readability indices, including Flesch Reading Ease (FRE), Flesch–Kincaid Grade Level (FKGL), Gunning Fog Index, Coleman-Liau Index, and Simple Measure of Gobbledygook (SMOG) Index, were used for this purpose.[12,20]

The FRE score ranges from 1 to 100, with higher scores indicating easier text in terms of readability. The FKGL, Gunning Fog Index, SMOG Index, and Coleman–Liau Index are tools that approximate the U.S. grade level needed for understanding a text.[21–24] Table 2 summarizes formulas[25] used for each index.

IBM SPSS Statistics for Windows, Version 25.0 (IBM Corp.) was used for the statistical analysis of the data. Descriptive statistics were used to report the variables of interest, with continuous variables being represented by their mean and standard deviation. Normality of data was evaluated using Shapiro–Wilk test. The Mann–Whitney test was used to compare the mean values for those without normal distribution, and independent *t* test was used for variables with a normal distribution. A *P* value of 0.05 or less was considered statistically significant.

Results

In this study, a total of 497 contents and responses including 71 contents from existing websites (9 from Healthline, 8 from NEI, 8 from National Health Service, 7 from WebMD, 10 from Prevent Blindness, 15 from Cleveland, 8 from Mayo Clinic, and 6 from AAO), 213 standard responses from ChatGPT-4, and 213 simplified responses from ChatGPT-4 were collected, uniformed, graded, and analyzed.

Assessment of Appropriateness

It was found that trained specialists preferred the standard responses provided by ChatGPT-4, and perceived these responses to be more comprehensive while maintaining accuracy compared with contents from the existing websites (Table 3).

The simplified ChatGPT-4 responses matched existing websites in terms of personal preference, accuracy, and comprehensiveness. However, they were preferred over WebMD and found to be more comprehensive compared with the contents of AAO and Prevent Blindness (Table 4).

Figure 1 shows a comparison of the average values of personal preference, comprehensiveness, and accuracy among the contents of existing websites and generated ChatGPT-4 responses.

Readability Assessments

The comparison of the aforementioned websites showed a range of readability, with some websites providing content that is easier to understand. Based on the average scores across various readability metrics, the NEI and WebMD emerged as easier to comprehend. National Eye Institute demonstrated lower FKGLs (7.13 ± 2.53) and higher FRE scores (62.76 ± 19.86), indicating that their content was understandable by a wider range of audience, including those with lower reading proficiency. WebMD followed closely with a FKGL of 7.05 ± 2.97 and a FRE score of 59.31 ± 24.32. On the other end of the spectrum, the Mayo Clinic and Healthline presented the most complex content, with FKGLs of 9.67 ± 2.29 and 8.86 ± 2.46, respectively (Figs 2 and 3). Standard ChatGPT-4 responses consistently required a reading level of grade 10 and higher according to the FKGL metric across all examined websites. The simplified ChatGPT-4 responses consistently exhibited a pattern where the required reading comprehension level, as determined by the FKGL metric, corresponded to sixth to seventh grade (Table 5).

Figure 3 illustrates the average grade level necessary for understanding content across 3 categories: all websites, standard ChatGPT-4 responses, and simplified ChatGPT-4 responses.

Discussion

Standard ChatGPT-4 responses on uveitis are preferred and are more comprehensive while being equally accurate compared with existing websites. Additionally, almost all simplified ChatGPT-4 responses maintain the same accuracy, comprehensiveness, and preference compared with the existing websites when evaluated by experts. These findings underscore the promising role of artifical intelligence (AI) in delivering health-care information.

In assessing readability, the study employed several metrics, each yielding unique insights due to their distinct focus areas and calculation methods. The Flesch–Kincaid Ease Score, FKGL, Gunning Fog Index, Coleman–Liau Index, and SMOG Index were used. The FKGL gauges readability based on word and sentence length, whereas the Coleman–Liau Index considers characters per word. The SMOG Index, focusing on complex word usage, often rates texts as more challenging, illustrating how different readability indices can yield varied assessments of the same content. The diversity in results from these readability indices highlights the multidimensional nature of text complexity. Such variability can be attributed to the distinct elements each index evaluates, such as sentence structure, word length, and syllable count. Applying all these measures ensured an accurate assessment of readability. Considering all these metrics, we found that standard ChatGPT-4 responses required a higher level of education to comprehend, whereas simplified ChatGPT-4 responses allowed a lower level of education compared with existing websites.

Table 2. Formula Used for Calculating Each Readability Index

Readability Index	Formula
Flesch Reading Ease (FRE) score[24]	206.835 − (1.015 × total words/total sentence) − (84.6 × total syllables/total word)
Flesch–Kincaid Grade Level[24]	(0.39 × total words/total sentence) + (11.8 × total syllables/total word) − 15.59
Gunning Fog index[25]	0.4 × (total words/total sentence + percentage of ≥ 3-syllable words)
Coleman–Liau Index	0.0588 × (characters/100 words) − 0.296 × (sentences/100 words) − 15.8
Simple Measure of Gobbledygook index (SMOG)[26]	3 + (square root of ≥ 3-syllable words count/30 sentences)

Table 3. Comparison of Personal Preference, Comprehensiveness, and Accuracy of Provided Contents among Existing Websites and Standard ChatGPT-4 Responses

	Healthline (Mean ± SD)	Standard ChatGPT-4 (Mean ± SD)	*P* Value
Personal preference	3.33 ± 0.40	4.59 ± 0.22	**0.0003**
Comprehensiveness	3.00 ± 1.00	4.78 ± 0.41	**0.0015**
Accuracy	4.00 ± 0.70	4.33 ± 0.50	0.2848
	NEI (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.45 ± 0.50	4.37 ± 0.21	**0.0007**
Comprehensiveness	2.88 ± 0.64	4.50 ± 0.53	**0.0011**
Accuracy	4.00 ± 0.53	4.25 ± 0.46	0.3317
	NHS (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.45 ± 0.68	4.50 ± 0.17	**0.0023**
Comprehensiveness	3.13 ± 1.24	4.75 ± 0.46	**0.0056**
Accuracy	4.13 ± 0.64	4.25 ± 0.46	0.6985
	WebMD (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	2.90 ± 0.46	4.38 ± 0.23	**0.0014**
Comprehensiveness	2.43 ± 0.97	4.57 ± 0.53	**0.0025**
Accuracy	4.14 ± 0.69	4.14 ± 0.37	0.9358
	Prevent Blindness (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.23 ± 0.78	4.43 ± 0.47	**0.0025**
Comprehensiveness	2.90 ± 0.56	4.40 ± 0.69	**0.0006**
Accuracy	4.20 ± 0.63	4.30 ± 0.48	0.7518
	Cleveland Clinic (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.37 ± 0.68	4.46 ± 0.27	**0.0001**
Comprehensiveness	2.67 ± 0.97	4.70 ± 0.45	**0.0001**
Accuracy	4.00 ± 0.66	4.27 ± 0.45	0.0655
	Mayo Clinic (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.04 ± 0.60	4.50 ± 0.17	**0.0006**
Comprehensiveness	3.00 ± 0.75	4.62 ± 0.51	**0.0015**
Accuracy	4.12 ± 0.64	4.38 ± 0.51	0.4237
	AAO (Mean ± SD)	**Standard ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	2.94 ± 0.44	4.50 ± 0.18	**0.0032**
Comprehensiveness	2.00 ± 0.63	4.67 ± 0.51	**0.0028**
Accuracy	4.16 ± 0.75	4.33 ± 0.51	0.7150

AAO = American Academy of Ophthalmology; NEI = National Eye Institute; NHS = National Health Service; SD = standard deviation.
Bold values are statistically significant.

Our results mirror a recent study by Mondal et al, in which they assessed the generated text by ChatGPT-4 for its applicability in educating patients about dermatological conditions. They concluded that ChatGPT-4 is capable of creating educational paragraphs that are easily comprehensible to individuals ranging from high school students to those who have just started college.[26]

The study conducted by Kianian et al focused on evaluating the effectiveness of ChatGPT-4 in generating appropriate and easy-to-read information about uveitis. They issued a general query to ChatGPT-4, prompting it to "write patient-targeted health information on uveitis" and then assessed the appropriateness and readability of the content produced. The evaluation involved grading the material as either appropriate or inappropriate, and the FKGL formula was utilized to measure readability. The results demonstrated that ChatGPT-4 consistently generated appropriate patient education materials in all 10 attempts, with adjustable levels of readability. These findings indicate that ChatGPT-4 has the potential to be a valuable tool for clinicians in creating patient-friendly health information regarding uveitis.[14] Our study differs from the one conducted by Kianian et al in several ways. Unlike their approach, the appropriateness in this study was assessed on various aspects, including the personal preferences of the

Table 4. Comparison of Personal Preference, Comprehensiveness, and Accuracy of Provided Contents among Existing Websites and *Simplified* ChatGPT-4 Responses

	Healthline (Mean ± SD)	Simplified ChatGPT-4 (Mean ± SD)	*P* Value
Personal preference	3.33 ± 0.40	3.62 ± 0.67	0.3215
Comprehensiveness	3.00 ± 1.00	3.44 ± 0.52	0.1652
Accuracy	4.00 ± 0.70	3.89 ± 0.33	0.6945
	NEI (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal p	3.45 ± 0.50	3.29 ± 0.48	0.4143
Comprehensiveness	2.88 ± 0.64	3.50 ± 0.53	0.0591
Accuracy	3.88 ± 0.35	3.75 ± 0.46	0.3317
	NHS (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.45 ± 0.68	3.37 ± 0.37	0.9148
Comprehensiveness	3.13 ± 1.24	3.12 ± 0.64	0.9099
Accuracy	4.13 ± 0.64	3.75 ± 0.88	0.3960
	WebMD (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	2.90 ± 0.46	3.47 ± 0.42	**0.0265**
Comprehensiveness	2.43 ± 0.97	3.28 ± 0.48	0.0628
Accuracy	3.71 ± 0.48	3.71 ± 0.48	0.2022
	Prevent Blindness (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.23 ± 0.78	3.40 ± 0.51	0.4909
Comprehensiveness	2.90 ± 0.56	3.80 ± 0.63	**0.0055**
Accuracy	3.60 ± 0.51	3.80 ± 0.63	0.1681
	Cleveland Clinic (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.37 ± 0.68	3.35 ± 0.47	0.2153
Comprehensiveness	2.67 ± 0.97	3.26 ± 0.70	0.0815
Accuracy	4.00 ± 0.66	3.53 ± 0.63	0.0655
	Mayo Clinic (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	3.04 ± 0.60	3.25 ± 0.34	0.5883
Comprehensiveness	3.00 ± 0.75	3.37 ± 0.74	0.3082
Accuracy	3.88 ± 0.64	3.62 ± 0.91	0.2438
	AAO (Mean ± SD)	**Simplified ChatGPT-4 (Mean ± SD)**	***P* value**
Personal preference	2.94 ± 0.44	3.38 ± 0.49	0.1387
Comprehensiveness	2.00 ± 0.63	3.33 ± 0.51	**0.0067**
Accuracy	4.16 ± 0.75	3.83 ± 0.40	0.3384

AAO = American Academy of Ophthalmology; NEI = National Eye Institute; NHS = National Health Service; SD = standard deviation.
Bold values are statistically significant.

graders, the comprehensiveness and accuracy of the materials generated in standard and simplified versions, and how these compared with existing websites in a masked manner. Additionally, our study aimed to simulate the patient experience by posing specific questions a patient might have about uveitis. Moreover, 5 readability indices were used to ensure the accuracy of the findings.

The objective of this study was to explore the potential of ChatGPT-4 in generating appropriate and comprehensible responses on prompts about uveitis, thereby enabling patients to access information efficiently without browsing multiple websites. Interestingly, standard ChatGPT-4 responses were preferred and more comprehensive while maintaining the same level of accuracy, although they had lower readability compared with existing websites. Meanwhile, simplified responses from ChatGPT-4 exhibited the same levels of personal preference, comprehensiveness, and accuracy but with enhanced readability.

Balancing readability with the appropriateness of content is crucial for patient education. Standard ChatGPT-4 responses were more detailed, requiring a higher educational

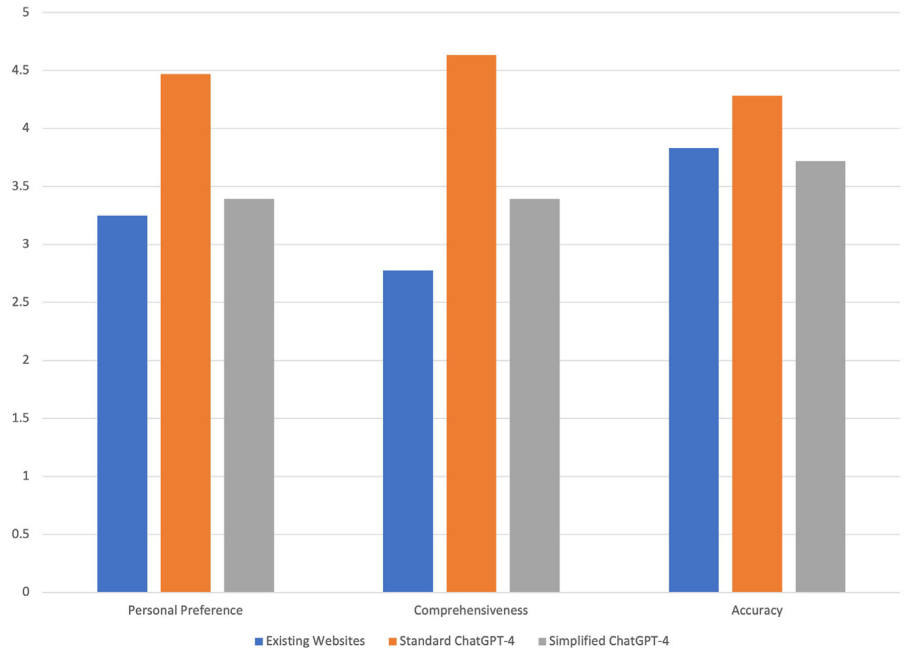

Figure 1. Comparison of appropriateness of contents of existing websites, standard ChatGPT-4 responses, and simplified ChatGPT-4 responses.

level (grade 10 and above). In contrast, simplified ChatGPT-4 responses were more accessible, suitable for a sixth- to seventh-grade level. The varied readability indicates the potential of ChatGPT-4 to tailor responses to varying literacy levels and customize them based on an individual's educational background.

Our study does have a few limitations. One limitation is that ChatGPT-4 requires a subscription, whereas the websites mentioned are free to access. Additionally, there are challenges related to trusting in AI and patient preference. However, this study represents a step toward building trust is AI by demonstrating that ChatGPT-4 has the potential to

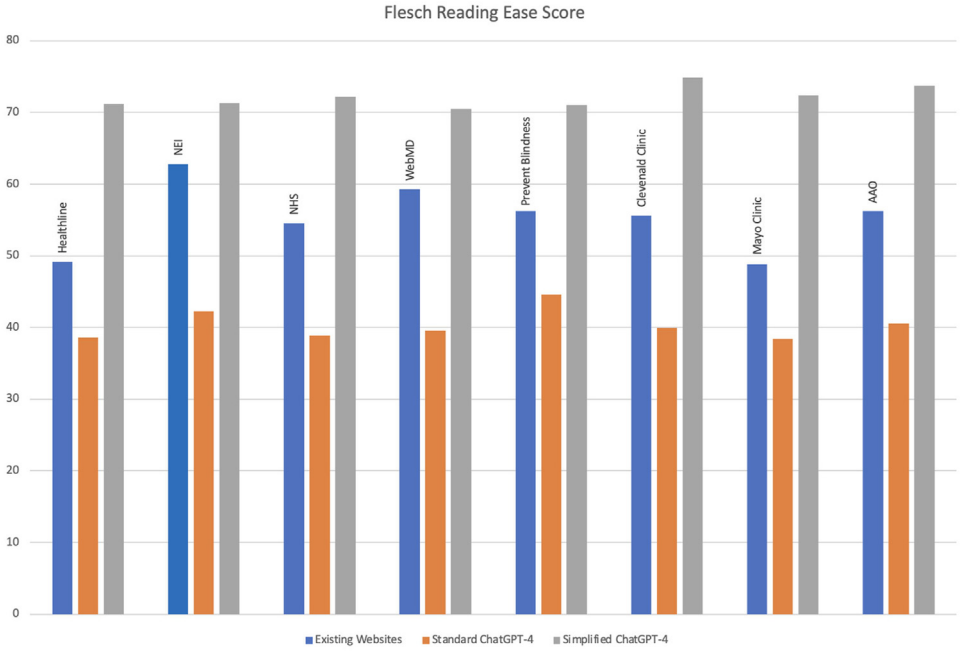

Figure 2. Flesch Reading Ease Scores of existing websites, standard ChatGPT-4 responses, and simplified ChatGPT-4 responses. AAO = American Academy of Ophthalmology; NEI = National Eye Institute; NHS = National Health Service.

Table 5. Comparison of Readability Indices among Existing Websites, Standard ChatGPT-4 Responses, and Simplified ChatGPT-4 Responses

	Healthline (Mean ± SD)	Standard ChatGPT (Mean ± SD)	Simplified ChatGPT (Mean ± SD)
Flesch–Kincaid Grade Level	8.86 ± 2.46	11.26 ± 1.74	6.64 ± 1.38
Flesch Reading Ease score	49.14 ± 16.06	38.55 ± 12.41	71.17 ± 7.58
Gunning Fog Index	12.32 ± 3.39	12.15 ± 2.70	8.40 ± 1.62
SMOG Index	11.51 ± 2.22	13.23 ± 1.80	9.73 ± 1.29
Coleman–Liau Index	10.58 ± 2.52	13.41 ± 1.53	8.23 ± 1.48
	NEI (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	7.13 ± 2.53	10.81 ± 1.45	7.03 ± 1.38
Flesch Reading Ease score	62.76 ± 19.86	42.28 ± 12.70	71.28 ± 6.38
Gunning Fog Index	8.66 ± 2.44	11.80 ± 2.19	9.11 ± 1.74
SMOG Index	9.82 ± 1.67	12.71 ± 1.59	9.90 ± 1.31
Coleman–Liau Index	8.18 ± 2.63	12.70 ± 1.61	7.82 ± 1.16
	NHS (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	8.52 ± 1.35	10.91 ± 1.95	6.46 ± 1.18
Flesch Reading Ease score	54.52 ± 7.08	38.80 ± 13.07	72.18 ± 6.08
Gunning Fog Index	11.61 ± 2.47	11.70 ± 3.18	8.17 ± 1.37
SMOG Index	11.68 ± 1.58	12.61 ± 2.11	9.53 ± 1.14
Coleman–Liau Index	10.18 ± 1.64	13.02 ± 1.66	7.61 ± 1.27
	WebMD (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	7.05 ± 2.97	10.71 ± 1.43	6.71 ± 1.74
Flesch Reading Ease score	59.31 ± 24.32	39.55 ± 13.19	70.51 ± 10.31
Gunning Fog Index	7.91 ± 2.16	11.37 ± 2.18	8.44 ± 1.96
SMOG Index	9.10 ± 1.42	12.38 ± 1.47	9.95 ± 1.76
Coleman–Liau Index	8.82 ± 2.74	12.97 ± 1.50	8.22 ± 1.60
	Prevent Blindness (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	8.29 ± 1.94	10.32 ± 1.46	6.86 ± 1.57
Flesch Reading Ease score	56.21 ± 16.55	44.56 ± 13.06	71.04 ± 7.92
Gunning Fog Index	9.55 ± 2.12	11.20 ± 1.94	8.61 ± 1.79
SMOG Index	10.74 ± 1.57	12.23 ± 1.38	9.85 ± 1.49
Coleman–Liau Index	9.20 ± 2.85	12.42 ± 1.88	8.36 ± 1.56
	Cleveland Clinic (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	7.90 ± 2.27	10.48 ± 1.90	6.08 ± 1.36
Flesch Reading Ease score	55.60 ± 14.96	39.88 ± 12.63	74.91 ± 5.80
Gunning Fog Index	9.40 ± 3.00	11.36 ± 2.40	7.93 ± 1.63
SMOG Index	10.14 ± 2.15	12.38 ± 1.74	9.18 ± 1.15
Coleman–Liau Index	10.11 ± 2.13	13.09 ± 2.00	7.33 ± 1.16
	Mayo Clinic (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	9.67 ± 2.29	10.25 ± 1.30	6.52 ± 1.36
Flesch Reading Ease score	48.81 ± 22.46	38.38 ± 9.78	72.32 ± 5.95
Gunning Fog Index	11.46 ± 2.34	10.41 ± 1.79	8.36 ± 1.55
SMOG Index	11.27 ± 1.65	11.63 ± 1.35	9.55 ± 1.14
Coleman–Liau Index	10.50 ± 2.90	12.22 ± 1.32	7.55 ± 1.07
	AAO (Mean ± SD)	**Standard ChatGPT (Mean ± SD)**	**Simplified ChatGPT (Mean ± SD)**
Flesch–Kincaid Grade Level	7.96 ± 1.55	10.86 ± 1.62	6.31 ± 1.54
Flesch Reading Ease score	56.21 ± 9.32	40.58 ± 13.37	73.71 ± 6.98
Gunning Fog Index	9.55 ± 2.72	11.43 ± 2.49	8.13 ± 1.78
SMOG Index	10.36 ± 1.61	12.66 ± 1.76	9.45 ± 1.43
Coleman–Liau Index	10.20 ± 1.61	12.96 ± 1.47	7.65 ± 1.26

AAO = American Academy of Ophthalmology; NEI = National Eye Institute; NHS = National Health Service; SD = standard deviation; SMOG = Simple Measure of Gobbledygook.

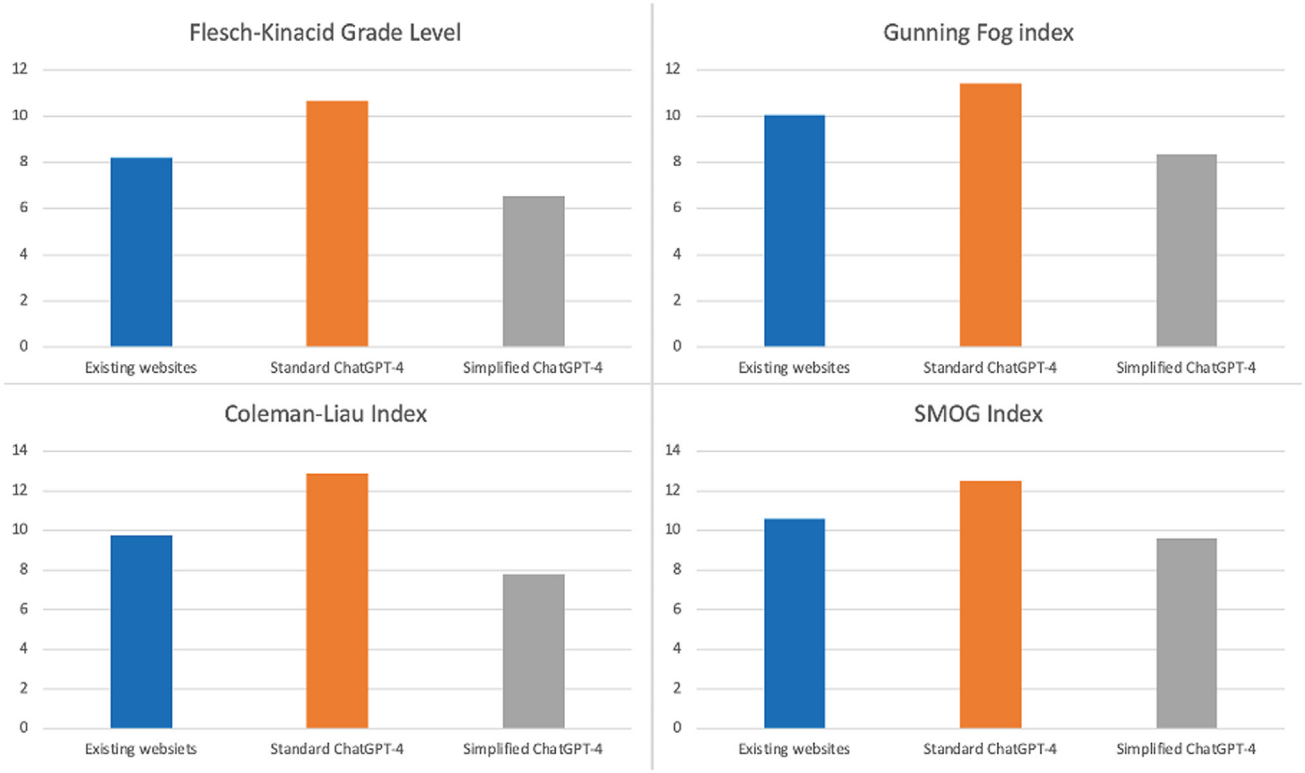

Figure 3. The average grade level necessary for understanding contents and responses across 3 categories: existing websites, standard ChatGPT-4 responses, and simplified ChatGPT-4 responses. SMOG = Simple Measure of Gobbledygook.

generate appropriate and comprehensible responses to patient queries on uveitis.

The findings of this study can have important implications for the integration of AI in patient education. The ability of ChatGPT-4 to adjust response complexity while maintaining information quality suggests its potential in creating customized responses to prompts about uveitis. Future research could focus on how AI can be effectively integrated into clinical settings, assess its impact on patient outcomes and satisfaction, and explore its role in addressing health literacy disparities.

Conclusion

ChatGPT-4 has the potential to generate accurate and comprehensive responses to queries about uveitis, which experts prefer over those from mentioned websites, although it requires a higher level of education to comprehend. Additionally, it can generate responses that are not only comparable with those websites but also more understandable. The results highlight the ability of ChatGPT-4 to customize responses about uveitis to suit each patient's comprehension level.

Footnotes and Disclosures

Originally received: February 5, 2024.
Final revision: July 22, 2024.
Accepted: August 5, 2024.
Available online: August 8, 2024. Manuscript no. XOPS-D-24-00037.

[1] Department of Ophthalmology, Byers Eye Institute, Stanford University, Palo Alto, California.

[2] Birat Aankha Aspatal, Biratnagar, Nepal.

[3] Department of Ophthalmology, Birat Medical College and Teaching Hospital, Kathmandu University, Biratnagar, Nepal.

[4] Dr. Shroff Charity Eye Hospital, New Delhi, India.

[5] National Healthgroup Eye Institute, Tan Tock Seng Hospital, Singapore.

[6] Gyeongsang National University Hospital, Jinju, South Korea.

[7] Bolor Melmii Eye Hospital, Ulaanbaatar, Mongolia.

[8] The Department of Ophthalmology, Faculty of Medicine, Tanta University, Tanta, Egypt.

[9] The Department of Ophthalmology, Viet Nam National Children's Hospital, Hanoi, Viet Nam.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s):

Q.D.N.: Grants – Acelyrin, Alumis, Belite Bio, Novartis, Priovant, Regeneron; Consultant – Bausch and Lomb, Regeneron; Participation on a Data Safety Monitoring Board or Advisory Board – Alvotech; Leadership or fiduciary role in other board society, committee or advocacy group, paid or unpaid – Tourmaline Scientific Advisory Board.

HUMAN SUBJECTS: No human subjects were included in this study. No animal subjects were used in this study.

Author Contributions:

Conception and design: Mohammadi, Khatri, Jain, Thng, Yoo, Yavari, Mobasserian, Nguyen

Data collection: Mohammadi, Khatri, Jain, Thng, Yoo, Yavari, Mobasserian, Akhavanrezayat, Elaraby, Ganbold, El Feky, Gupta, Hung, Nguyen

Analysis and interpretation: Mohammadi, Yoo, Mobasserian, Akhavanrezayat, Elaraby, Ganbold, El Feky, Gupta, Hung, Nguyen

Obtained funding: Study was performed as part of regular employment duties at Stanford University. No additional funding was provided for this work.

Overall responsibility: Mohammadi, Khatri, Jain, Thng, Yavari, Akhavanrezayat, Elaraby, Ganbold, El Feky, Gupta, Hung, Nguyen

Abbreviations and Acronyms:
AAO = American Academy of Ophthalmology; **FKGL** = Flesch−Kincaid Grade Level; **FRE** = Flesch Reading Ease; **NEI** = National Eye Institute; **SMOG** = Simple Measure of Gobbledygook.

Keywords:
Artificial intelligence, ChatGPT, Patient education, Uveitis.

Correspondence:
Quan Dong Nguyen, MD, MSc, Spencer Center for Vision Research, Byers Eye Institute at Stanford University, 2370 Watson Court, Suite 200, Palo Alto, CA 94303. E-mail: ndquan@stanford.edu.

References

1. Siddique SS, Suelves AM, Baheti U, Foster CS. Glaucoma and uveitis. *Surv Ophthalmol*. 2013;58:1−10.
2. Foster CS, Kothari S, Anesi SD, et al. The Ocular Immunology and Uveitis Foundation preferred practice patterns of uveitis management. *Surv Ophthalmol*. 2016;61:1−17.
3. Foster CS, Vitale AT. *Diagnosis & Treatment of Uveitis*. New Delhi, India: JP Medical Ltd; 2013.
4. Roundtable on Health Literacy, Board on Population Health and Public Health Practice. *Institute of Medicine*. Health Literacy: Improving Health, Health Systems, and Health Policy Around the World: Workshop Summary. Washington (DC): National Academies Press (US); 2013. http://www.ncbi.nlm.nih.gov/books/NBK202438/. Accessed November 23, 2023.
5. Matsuoka S, Tsuchihashi-Makaya M, Kayane T, et al. Health literacy is independently associated with self-care behavior in patients with heart failure. *Patient Educ Couns*. 2016;99:1026−1032.
6. Aaby A, Friis K, Christensen B, et al. Health literacy is associated with health behaviour and self-reported health: a large population-based study in individuals with cardiovascular disease. *Eur J Prev Cardiol*. 2017;24:1880−1888.
7. Balakrishnan MP, Herndon JB, Zhang J, et al. The association of health literacy with preventable emergency department visits: a cross-sectional study. *Acad Emerg Med*. 2017;24:1042−1050.
8. Mueller CM, Ward L, O'Keefe GAD. Health literacy, medication adherence, and quality of life in patients with uveitis. *Ocul Immunol Inflamm*. 2022;30:1060−1067.
9. Papagiannuli E, Edmunds MR, Scollo P, et al. Do demographic factors influence uveitis patients' understanding of uveitis? *Ocul Immunol Inflamm*. 2017;25:790−796.
10. Muir KW, Lee PP. Health literacy and ophthalmic patient education. *Surv Ophthalmol*. 2010;55:454−459.
11. Kickbusch I, Pelikan JM, Apfel F, Tsouros A. Health literacy. WHO regional office for europe. https://iris.who.int/handle/10665/128703; 2013. Accessed November 24, 2023.
12. Mikhail D, Visscher KL, Chen N, et al. Patient-appropriate health literacy educational materials in ophthalmology. *Can J Ophthalmol*. 2015;50:19−25.
13. Stossel LM, Segar N, Gliatto P, et al. Readability of patient education materials available at the point of care. *J Gen Intern Med*. 2012;27:1165−1170.
14. Kianian R, Sun D, Crowell EL, Tsui E. The use of large language models to generate education materials about uveitis. *Ophthalmol Retina*. 2023;0.
15. Vaishya R, Misra A, Vaish A. ChatGPT: is this version good for healthcare and research? *Diabetes Metab Syndr*. 2023;17:102744.
16. 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.
17. Rojas-Carabali W, Sen A, Agarwal A, et al. Chatbots vs. Human experts: evaluating diagnostic performance of chatbots in uveitis and the perspectives on AI adoption in ophthalmology. *Ocul Immunol Inflamm*. 2023:1−8.
18. 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;101:829−831.
19. Johnson SB, King AJ, Warner EL, et al. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. *JNCI Cancer Spectr*. 2023;7:pkad015.
20. Patel AJ, Kloosterboer A, Yannuzzi NA, et al. Evaluation of the content, quality, and readability of patient accessible online resources regarding cataracts. *Semin Ophthalmol*. 2021;36:384−391.
21. Kincaid JP, Fishburne RP, Rogers RL, Chissom BS. *Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel*. Millington, TN: Naval Technical Training Command; 1975.
22. Gunning R. *The Technique of Clear Writing*. New York, NY: McGraw-Hill; 1952.
23. McLaughlin GH. SMOG grading−a new readability formula. *J Read*. 1969;12:639−646.
24. Coleman M, Liau TL. A computer readability formula designed for machine scoring. *J Appl Psychol*. 1975;60:283.
25. Yeung AWK, Goto TK, Leung WK. Readability of the 100 most-cited neuroimaging papers assessed by common readability formulae. *Front Hum Neurosci*. 2018;12:308.
26. Mondal H, Mondal S, Podder I. Using ChatGPT for writing articles for patients' education for dermatological diseases: a pilot study. *Indian Dermatol Online J*. 2023;14:482.