

Original Paper

# Evaluating ChatGPT's Utility in Biologic Therapy for Systemic Lupus Erythematosus: Comparative Study of ChatGPT and Google Web Search

Kai Li<sup>1</sup>, PhD; Yunfei Peng<sup>1\*</sup>, BA; Luyi Li<sup>1\*</sup>, BA; Bo Liu<sup>2</sup>, MM; Zhijian Huang<sup>2</sup>, MM

<sup>1</sup>School of Journalism and Communication, Guangxi University, Nanning, China

<sup>2</sup>Department of Rheumatology and Immunology, The Second Affiliated Hospital of Guangxi Medical University, Nanning, China

\*these authors contributed equally

## Corresponding Author:

Kai Li, PhD  
School of Journalism and Communication  
Guangxi University  
No.100, Daxue East Road  
Nanning 530004  
China  
Phone: 86 13367611322  
Email: [20160269@gxu.edu.cn](mailto:20160269@gxu.edu.cn)

## Abstract

**Background:** Systemic lupus erythematosus (SLE) is a life-threatening, multisystem autoimmune disease. Biologic therapy is a promising treatment for SLE. However, public understanding of this therapy is still insufficient, and the quality of related information on the internet varies, which affects patients' acceptance of this treatment. The effectiveness of artificial intelligence technologies, such as ChatGPT (OpenAI), in knowledge dissemination within the health care field has attracted significant attention. Research on ChatGPT's utility in answering questions regarding biologic therapy for SLE could promote the dissemination of this treatment.

**Objective:** This study aimed to evaluate ChatGPT's utility as a tool for users to obtain health information about biologic therapy for SLE.

**Methods:** This study extracted 20 common questions related to biologic therapy for SLE, their corresponding answers, and the sources of these answers from both Google Web Search and ChatGPT-4o (OpenAI). Then, based on Rothwell's classification, the questions were categorized into 3 main types: fact, policy, and value. The sources of the answers were classified into 5 categories: commercial, academic, medical practice, government, and social media. The accuracy and completeness of the answers were assessed using Likert scales. The readability of the answers was evaluated using the Flesch Reading Ease and Flesch-Kincaid Grade Level (FKGL) scores.

**Results:** The study found that, in terms of question types, ChatGPT-4o had the highest proportion of fact questions (10/20), followed by policy (7/20) and value (3/20). Google Web Search had the highest proportion of fact questions (12/20), followed by value (5/20) and policy (3/20). In terms of website sources, ChatGPT-4o's answers were sourced from 48 sources, with the majority coming from academic sources (29/48). Google Web Search provided answers from 20 sources, with an even distribution across all 5 categories. For accuracy, ChatGPT-4o's mean score of 5.83 (SD 0.49) was higher than that of Google Web Search (mean 4.75, SD 0.94), with a mean difference of 1.08 (95% CI 0.61-1.54). For completeness, ChatGPT-4o's mean score of 2.88 (SD 0.32) was higher than that of Google Web Search (mean 1.68, SD 0.69), with a mean difference of 1.2 (95% CI 0.96-1.44). For readability, the Flesch Reading Ease and Flesch-Kincaid Grade Level scores for ChatGPT-4o and Google Web Search were 11.7 and 14.9, and 16.2 and 20, respectively, indicating that both texts were of high reading difficulty, requiring readers to have a college graduate-level reading proficiency. When asking ChatGPT to respond at a sixth-grade level, the readability of the answers significantly improved.

**Conclusion:** ChatGPT's answers are characterized by accuracy, rigor, comprehensiveness, and professional supporting materials, and demonstrate humanistic care. However, the readability of the provided text is low, requiring users to have a college education background. Given the study's limitations in question scope, comparison dimensions, research perspectives, and language types, further in-depth comparative research is recommended.

**Keywords:** biologic therapy; SLE; ChatGPT; google web search; health information; systemic lupus erythematosus

## Introduction

ChatGPT (OpenAI) has gained significant attention for its robust capabilities in content generation since its launch. It can respond to user inquiries by providing relevant, accurate, and evidence-based information in a logically coherent and easily understandable manner [1]. ChatGPT-4o (OpenAI) and its subsequent versions have more natural human-computer interaction capabilities, enabling them to respond to user commands more efficiently [2]. Although ChatGPT was originally developed neither for the health care domain [3] nor explicitly for answering medical questions [4], its content generation potential in the health care field is particularly noteworthy. Studies have found that ChatGPT plays a positive role in helping users gain health knowledge and answering medical inquiries [5,6].

Ayik et al [7] and Javaid et al [8] have indicated that ChatGPT can assist users in answering common questions in the health care domain. Temsah et al [9] noted that ChatGPT offers high-quality health care information, thereby bridging the gap in health care access. According to Ruksakulpiwat et al [10], the medical information provided by ChatGPT can facilitate communication between patients and health care providers, enhance patients' self-care abilities, and improve treatment outcomes. Ayoub et al [5] are optimistic about ChatGPT's potential for knowledge acquisition in the health care domain, suggesting that in certain situations, ChatGPT could enhance or even replace the commonly used Google Web Search, becoming an important information source for patients.

Systemic lupus erythematosus (SLE) is a multisystem autoimmune disease with a prevalence that varies across populations, with a global average prevalence ranging from 12 to 39 per 100,000 individuals [11]. China ranks among the highest in the world in terms of the burden of SLE [12], with a prevalence ranging from 30.13 to 70.41 per 100,000 individuals. The disease is most commonly observed in women of reproductive age, particularly those between 20 and 40 years old [11]. SLE leads to severe organ complications and may even be life-threatening [13], requiring urgent medical attention. In recent years, with in-depth research on the pathogenesis of SLE, targeted therapy has been proposed and has gradually gained attention. Some new biologics have been developed and introduced into clinical use. According to the 2023 annual "Patient Chart Dynamix: SLE (US)" and "Patient Chart Dynamix: SLE (EU5)" published by Spherix Global Insights, an increasing number of patients with SLE in the United States and 5 European countries are receiving biologic therapy, with the proportion of patients receiving such treatment rising by 32% compared with 2022 [14]. Currently, belimumab and anifrolumab have been successfully approved by the Food and Drug Administration for the treatment of SLE [15]. In China, belimumab, telitacicept, and

rituximab are the 3 biologic agents currently used in clinical practice [16].

Although biologic agents play a crucial role in SLE treatment, Liu et al [17] found that many patients with SLE lack knowledge about treatment options and disease management. This not only significantly affects their attitudes, treatment practices, and willingness to choose biologic therapy but also hinders their active participation in disease management, ultimately impacting their prognosis [17]. A study by Waldron et al [18] found that when health care providers fail to provide clear and consistent information during diagnosis, patients with SLE experience distress and confusion. In such situations, patients with SLE are highly likely to use the internet to meet their information needs [19]. According to the data from the Pew Research Center, 77% of health information seekers reported that their search began with applications such as Google, Bing, or Yahoo [20]. Jacquemart and Zweigenbaum [21] found that Google is the most effective search engine for retrieving answers to clinical questions. Lopes and Ribeiro [22] also found that Google significantly outperforms other search engines in accessing health information. Nowadays, with the emergence of large language models (LLMs), such as ChatGPT, some users are shifting from traditional web searches to these models as a primary source of health care information [5,6,23]. Traditional search engines, such as Google Bard and Microsoft Copilot, have also integrated LLMs into their platforms [24]. Hopkins et al [25] believed that LLMs (eg, ChatGPT) will fundamentally transform the way patients seek health care information.

The readability of health care texts directly impacts how effectively readers acquire health information from them. According to the readability assessment standards set by the National Institutes of Health and the American Medical Association, presenting health information on websites at a sixth-grade reading level or lower may help improve patients' understanding of their condition and the treatment plans recommended by health care providers [26]. A study by Daraz et al [27] in the United States and Canada found that the vast majority of health information on the internet has poor readability, making it unsuitable for the general public. Furthermore, it may lead to misinformation and have a detrimental effect on health. Massie et al [28] noted that the readability of the American Society of Metabolic Surgery's Patient Information Publications is poor and does not meet the standard of reading levels at or below the sixth grade. However, a study by Moons and Van Bulck [29] found that ChatGPT is a valuable tool for improving the readability and comprehensibility of health information. It can generate health care information suitable for the general public and easy to understand, based on user instructions, thereby addressing the issue of poor readability in health information on the internet [29].

Previous studies have evaluated the effectiveness of ChatGPT-3.5 (OpenAI) in responding to SLE-related questions and found that, in response to certain questions, its answers were not comprehensive or accurate and were sometimes unrelated to the questions [30]. ChatGPT-4o is widely recognized as a relatively new artificial intelligence model with a high user adoption rate [31]. However, no studies have yet evaluated the effectiveness of the new version, ChatGPT-4o, in responding to common SLE-related questions. Biologic therapy is a relatively new and important treatment for SLE and is regarded with high expectations. However, public awareness of this therapeutic approach remains insufficient [17]. Furthermore, the quality of SLE-related information on the internet varies, with some content containing advertisements and misleading information. This significantly hinders patients' acceptance of biologic therapy [32], ultimately causing a significant negative impact on the clinical outcomes of patients with SLE. Therefore, this study aims to replicate users' search behavior on Google Web Search to evaluate the utility of ChatGPT-4o in responding to common questions about biologic therapy for SLE. Specifically, we compare the following three aspects: (1) the most common questions related to biologic therapy for SLE, categorized by question types and topics; (2) the answers to these questions; and (3) the sources of these answers.

## Methods

### Overview

This study extracted 20 common questions related to biologic therapy for SLE, their corresponding answers, and the sources of these answers from both Google Web Search and ChatGPT-4o [33]. First, this study used the "People also ask" function in Google Web Search by entering the search terms "biologic therapy" and "systemic lupus erythematosus." We selected the 20 most common questions about biologic therapy for SLE and recorded their answers and sources. Second, we instructed ChatGPT-4o to execute the command "Do a Google search with the search term 'biologic therapy,' 'systemic lupus erythematosus,' and record 20 common questions related to the search term." Similarly, we recorded the questions, their answers, and the sources of those answers. Third, the obtained questions and the sources of answers were classified. Fourth, the accuracy, completeness, and readability of the answers were evaluated. In total, 2 rheumatology physicians (BL and ZH) were involved throughout the process of question selection and answer evaluation. In the event of a disagreement, a third physician was included in the discussion to reach a consensus.

### Google Web Search Operation

On July 17, 2024, a new installation of the Google Chrome browser was performed, and all browsing history, cookies, site data, cached images, files, sponsored websites, and advertisements were cleared before initiating the search. The keywords "biologic therapy" and "systemic lupus erythematosus" were entered simultaneously, and questions were

identified using the "People Also Ask" function in Google Web Search. Additional related questions were obtained by clicking on each question.

We conducted 10 rounds of searches. During the selection of relevant questions, we found that once the sample size exceeded 30, the system started generating irrelevant or duplicate questions. Therefore, to ensure a representative sample, we selected only the first 30 questions from each round, yielding a total of 300 questions. After merging similar questions and analyzing their frequencies, 2 professional physicians (BL and ZH) assigned importance levels to each question according to their clinical experience. Finally, 20 questions related to biologic therapy for SLE were selected, along with their corresponding answers and the website information sources.

### ChatGPT Operation Execution

The following command was entered into ChatGPT-4o: "Do a Google search with the search term 'biologic therapy,' 'systemic lupus erythematosus,' and record 30 common questions related to the search term." Then, each question was individually entered into ChatGPT-4o, and the corresponding answers and their sources were recorded. We conducted 10 rounds of searches, selecting 30 questions per round, which resulted in a total of 300 questions. We then used the same selection process for questions with Google Web Search, ultimately extracting 20 questions related to biologic therapy for SLE, along with their corresponding answers and website information sources.

Notably, unlike Google Web Search, ChatGPT-4o provided 2 or more sources for its answers. We used a 2-step verification process to assess whether hallucinations exist in its responses: First, each link attached to the answers was manually accessed and verified to check whether the link was valid and if the page content was consistent with the information provided in the answer. If the link was invalid or the page content was irrelevant to the answer, it was considered a potential indicator of hallucination. Second, based on the accuracy ratings given by 2 professional physicians (BL and ZH; refer to Accuracy and Completeness Evaluation section), if the average score of any answer was below 3, it was considered a hallucination, indicating that the answer contained factual errors or that the main information could not be substantiated [34]. No hallucination was found in the links or answers provided by ChatGPT after the 2-step verification. After removing duplicate links, we compiled statistics on the final website information sources.

### Inclusion and Exclusion Criteria

This study invited 2 professional physicians to screen and extract questions that are highly relevant to the topic and of significant concern to patients based on their clinical experience. The selected questions and answers were required to be closely related to biologic therapy for SLE. The inclusion criteria involved any question related to the terms "biologic therapy," "biological therapy," "biologics therapy," "biologic treatment," "rituximab," "belimumab," "anifrolumab," and "systemic lupus erythematosus," "lupus," or

“SLE.” The exclusion criteria included duplicate questions and questions unrelated to biologic therapy for SLE (eg, “What is the enemy of lupus?”).

## Ethical Considerations

This study analyzes publicly available data from Google Web Search and ChatGPT-4o. No personal or identifiable patient data were used. According to the Declaration of Helsinki established by the World Medical Association [35], this study does not involve human participants or personal data. Therefore, it is exempt from institutional review board review.

## Comparison Analysis

The common questions related to biologic therapy for SLE in this study were categorized using a modification of the

Rothwell system [36]. According to this system, questions can be classified into 3 main types (fact, policy, and value), as well as 10 subtypes, that are education, accessibility, timeline of effectiveness, technical details, cost, indications or management, risks or complications, future prospects and challenges, longevity, and evaluation of treatment (refer to Table 1 for details). The website information sources were categorized into 5 types—commercial, academic, medical practice, government, and social media [33,37] (refer to Table 2 for details). Furthermore, 2 rheumatology physicians (BL and ZH) conducted question classification. In the event of disagreements, a third physician was brought in to achieve a consensus through discussion.

**Table 1.** Question types categorized by Rothwell’s classification [36]. The questions are categorized into 3 main types and subsequent subtypes.

Categorization	Explanation
Fact	Asks whether something is true and to what extent, objective information
Education	Educational information about biologic therapies for SLE <sup>a</sup> , not fitting into other subcategories
Accessibility	Where and how patients can get biologic treatment for SLE
Timeline of effectiveness	Length of time for the biologic therapy to show effectiveness in SLE, rather than just recovery milestones
Technical details	Details of how biologic therapies work for SLE, including dosing, method of delivery, and so on
Cost	Cost of treatment or rehabilitation
Policy	Asks whether a specific course of action should be taken to solve a problem
Indications or management	Therapy indications and alternatives, timing of biologic therapy for SLE, and posttreatment monitoring and management
Risks and complications	Risks and complications before, during, or after biologic treatment, including the rehabilitation period for SLE
Future prospects and challenges	The progress and difficulty in the development of biologic therapy for SLE
Value	Asks for the evaluation of an idea, object, or event
Longevity	Longevity and sustained effect of biologic therapy for patients with SLE
Evaluation of treatment	Successfulness, seriousness, or invasiveness of biologic therapy for SLE

<sup>a</sup>SLE: systemic lupus erythematosus.

**Table 2.** Categorization of website information sources.

Categorization	Explanation
Commercial	Organizations that provide public health information, including medical device, manufacturing, or pharmaceutical companies and news outlets
Academic	Universities, academic medical centers, or academic societies
Medical practice	Local hospitals or medical groups without clear academic affiliation
Government	Websites maintained by a national government
Social media	Blog, internet forums, support groups, and nonmedical organizations designed for information and video sharing

## Accuracy and Completeness Evaluation

The accuracy and completeness of responses generated by Google Web Search and ChatGPT-4o were evaluated by 2 professional physicians based on the guidelines published by the American College of Rheumatology [38]. Accuracy was assessed using a 6-point Likert scale (with 1=completely incorrect, 2=more incorrect than correct, 3=approximately equal correct and incorrect, 4=more correct than incorrect, 5=nearly entirely correct, and 6=completely correct).

Completeness was evaluated using a 3-point Likert scale: 1 indicated “incomplete” (the response addressed some aspects of the question but omitted important information), 2 indicated “adequate” (the response addressed all aspects of the question with sufficient information for completeness), and 3 indicated “comprehensive” (the response addressed all aspects and provided additional information or context beyond expectations) [39]. Statistical analysis was performed using IBM SPSS Statistics 27, using descriptive statistics



(mean and SD) and the Cohen  $\kappa$  test to assess the interrater reliability.

Readability Measurement

Readability is defined as the ease of understanding or comprehension attributed to the writing style. Although many different readability tests exist, their measurement basically relies on two elements: (1) vocabulary and (2) word or sentence complexity (ie, the number of characters per space, syllables per word, or words per sentence) [40]. This study used the assessment tool “Readable” to evaluate the readability of the texts generated by Google Web Search and ChatGPT-4o [41]. Readability was assessed using the Flesch Reading Ease (FRE) and Flesch-Kincaid Grade Level (FKGL), both of which are commonly used to evaluate the readability of health-related literature on the internet [27,28]. The FRE score ranges from 0 to 100, with lower scores

indicating higher reading difficulty [42]. The FKGL ranges from 0 to 18, representing the number of years of education required to comprehend the text. A lower score indicates easier readability [43].

Results

Question Analysis

The classification results for the 20 most common questions obtained from Google Web Search and ChatGPT-4o are presented in Table 3. Regarding the distribution of main question types, questions categorized under fact accounted for 12 from Google Web Search and 10 from ChatGPT-4o, those under policy for 3 and 7, and those under value for 5 and 3, respectively. Fact-category questions accounted for the highest proportion in both sources.

**Table 3.** 20 most common questions for biologic therapy for systemic lupus erythematosus from Google Web Search and ChatGPT-4o. The main types and subtypes of questions, categorized by Rothwell’s classification [36], are shown in parentheses.

20 most common questions from Google Web Search		20 most common questions from ChatGPT-4o	
1	What is the safest biologic for lupus? (Value, evaluation of treatment)	1	What is biologic therapy for SLE <sup>a</sup> ? (Fact, education)
2	What is the new biologic for lupus? (Fact, education)	2	How do biologic therapies work in treating SLE? (Fact, technical details)
3	Is rituximab good for lupus? (Value, evaluation of treatment)	3	What are the most common biologic therapies used for SLE? (Fact, education)
4	What is the new drug for lupus approved by the FDA <sup>b</sup> ? (Fact, accessibility)	4	What are the potential side effects of biologic therapies for SLE? (Policy, risks, and complications)
5	Can biologics cause lupus? (Policy, risks, and complications)	5	Are there any new biologic therapies being developed for SLE? (Fact, future prospects, and challenges)
6	How long does it take rituximab to work for lupus? (Fact, timeline of Effectiveness)	6	How does biologic therapy compare to traditional treatments for SLE? (Value, evaluation of treatment)
7	How effective is rituximab for lupus?(Value, evaluation of treatment)	7	Can biologic therapies be used in combination with other SLE treatments? (Policy, indications, or management)
8	What is the biologic treatment for systemic lupus erythematosus? (Fact, education)	8	How is the effectiveness of biologic therapies in SLE measured in clinical trials? (Fact, technical details)
9	Is rituximab FDA approved for SLE? (Fact, accessibility)	9	What are the long-term outcomes for SLE patients on biologic therapies? (Value, longevity)
10	What is the new treatment for lupus in the UK? (Fact, accessibility)	10	What is the cost and insurance coverage situation for biologic therapies for SLE? (Fact, cost)
11	Can rituximab cure lupus? (Value, evaluation of treatment)	11	How long does it take to see improvements with biologic therapy in lupus? (Fact, timeline of effectiveness)
12	How does rituximab work for SLE? (Fact, technical details)	12	What are the criteria for a patient with SLE to be eligible for biologic therapy? (Policy, indications, or management)
13	Is rituximab used to treat SLE? (Policy, indications, or management)	13	How do biologic therapies impact the overall quality of life for SLE patients? (Value, evaluation of treatment)
14	How do biologics help lupus? (Fact, technical details)	14	How often are biologic therapies administered for SLE patients? (Fact, technical details)
15	What is the best infusion for lupus? (Value, evaluation of treatment)	15	What is the role of Benlysta (belimumab) in SLE treatment? (Fact, technical details)
16	What is the self-injection for lupus? (Fact, technical details)	16	Are there any dietary or lifestyle changes recommended for SLE patients on biologic therapies? (Policy, indications, or management)
17	How often is Rituxan given for lupus?(Fact, technical details)	17	What are the approved biologic therapies for SLE? (Fact, education)
18	What are the indications for rituximab in SLE? (Policy, indications or management)	18	How are patients monitored while on biologic treatments for SLE? (Policy, indications, or management)

	20 most common questions from Google Web Search	20 most common questions from ChatGPT-4o
19	When will anifrolumab be FDA approved?(Fact, accessibility)	Are there any specific biologic therapies recommended for lupus nephritis? (Policy, indications or management)
20	Can adalimumab cause lupus? (Fact, risks and complications)	Can biologic therapies be used during pregnancy in SLE patients? (Policy, indications or management)

<sup>a</sup>SLE: systemic lupus erythematosus.

<sup>b</sup>FDA: Food and Drug Administration.

The distribution of question subtypes is illustrated in [Multimedia Appendix 1](#). Questions from Google Web Search fell into 7 subtypes, with the highest proportions in the evaluation of treatment (n=5), followed by accessibility (n=4) and technical details (n=4). Other subtypes included education (n=2), risks and complications (n=2), indications or management (n=2), and timeline of effectiveness (n=1). Questions generated by ChatGPT-4o covered 9 subtypes, with indications or management (n=6) having the highest proportion, followed by technical details (n=4), education (n=3), evaluation of treatment (n=2). The remaining subtypes, including timeline of effectiveness, cost, risks and complications, future prospects and challenges, and longevity, were each represented by 1 question.

## Answer Analysis

[Multimedia Appendix 2](#) presents the responses generated by ChatGPT-4o and Google Web Search. In terms of accuracy, the accuracy scores indicate that ChatGPT-4o achieved a higher mean score (mean 5.83, SD 0.49) compared with

Google Web Search responses (mean 4.75, SD 0.94), with a mean difference of 1.08 (95% CI 0.61-1.54). The ratings provided by the 2 physicians demonstrated good interrater reliability, with Cohen  $\kappa$  values of 0.72 for ChatGPT-4o and 0.78 for Google Web Search.

For completeness, ChatGPT-4o's mean score was 2.88 (SD 0.32), significantly higher than Google Web Search's 1.68 (SD 0.69), with a mean difference of 1.2 (95% CI 0.96-1.44). The ratings also showed strong interrater reliability, with Cohen  $\kappa$  values of 0.94 for ChatGPT-4o and 0.88 for Google Web Search.

In terms of readability, the FRE and FKGL scores for Google Web Search and ChatGPT-4o were 16.2 and 20, and 11.7 and 14.9, respectively ([Table 4](#)). These results indicate that both texts have a high level of reading difficulty, requiring readers to have a college graduate-level reading proficiency [41]. In terms of response length, answers from Google Web Search were more concise, while those from ChatGPT-4o were more detailed and comprehensive.

**Table 4.** Accuracy, completeness, and readability of Google Web Search and ChatGPT-4o answers on biologic therapy for systemic lupus erythematosus.

	Item	Google Web Search	ChatGPT-4o
1	Accuracy, mean (SD)	5.83 (0.49)	4.75 (0.94)
2	Completeness, mean (SD)	2.88 (0.32)	1.68 (0.69)
3	Readability		
	FKGL <sup>a</sup>	16.2	11.7
	FRE <sup>b</sup>	20	17.3

<sup>a</sup>FKGL: Flesch-Kincaid Grade Level.

<sup>b</sup>FRE: Flesch Reading Ease.

Furthermore, 3 of the 20 questions obtained from each source were similar ([Table 5](#)), but their answers differed significantly. The answers provided by Google Web Search were not comprehensive enough, lacking explanations of drug

mechanisms and missing information on new medications (eg, Saphnelo) and emerging therapies (eg, chimeric antigen receptor-T cell therapy).

**Table 5.** Similar questions on biologic therapy for systemic lupus erythematosus of Google web search and ChatGPT-4o.

Question	Google Web Search	ChatGPT-4o
1	What is the new biologic for lupus?	Are there any new biologic therapies being developed for SLE <sup>a</sup> ?
2	What is the biologic treatment for systemic lupus erythematosus?	What is biologic therapy for SLE?
3	How do biologics help lupus?	How do biologic therapies work in treating SLE?

<sup>a</sup>SLE: systemic lupus erythematosus.

## Analysis of Website Information Source

The distribution of website information sources differed notably between Google Web Search and ChatGPT-4o (Multimedia Appendix 3). Google Web Search provided answers from 20 sources: 2 commercial, 4 academic, 5 medical practice, 4 government, and 5 social media. ChatGPT-4o's answers were cited from 48 sources, including 9 commercial, 29 academic, 1 medical practice, 2 government, and 7 social media sources. Notably, ChatGPT-4o relied more heavily on academic sources, whereas the source types of Google Web Search were more evenly distributed.

## Discussion

### Principal Findings

This study aimed to evaluate the utility of ChatGPT in providing users with health information on biologic therapy for SLE. The key findings are described further.

### Accuracy and Rigor of Questions and Answers Provided by ChatGPT

On the one hand, the study found that ChatGPT-4o provided answers with higher accuracy than Google Web Search, which is consistent with the findings of Cohen et al [44]. On the other hand, among the 20 questions provided by Google Web Search, 9 were related to rituximab treatment for SLE, covering topics such as its accessibility, technical details, indications, and treatment evaluation. In contrast, the questions provided by ChatGPT-4o did not focus on rituximab treatment for SLE. There are significant differences between the questions regarding biologic therapy for SLE, which are of common concern to users, presented by both platforms. This finding is consistent with the findings of Dubin et al [33]. The reasons could be that Google provides an effective marketing channel for some pharmaceutical and health websites, enabling them to gain more exposure through bidding. Sponsored links are prioritized and displayed at the top or in prominent positions on the search results page [45-47]. However, many of these links may not only lack accurate clinical evidence but may also contain concerning misinformation [48]. In addition, search engine optimization techniques, through the optimization of website elements and the building of high-quality interlinks, effectively enhance a website's ranking and visibility in search engines, making it a key channel for increasing organic traffic [49]. However, this could also mean that high-quality, unoptimized medical content pages may be pushed down in rankings, replaced by lower-quality pages [50]. ChatGPT-4o did not generate questions related to rituximab treatment for SLE. This may be because although textbooks such as "Harrison's Principles of Internal Medicine" [51] and clinical guidelines such as "American College of Rheumatology Guidelines for Screening" [38] recommend the use of rituximab for treating SLE and its clinical application is widespread, rituximab has never been officially approved for SLE treatment at the legal level, nor is SLE listed as an indication in the drug's prescribing information. The "Usage Policies" published by

OpenAI under "Building with the OpenAI API Platform" explicitly state that ChatGPT-4o must not provide tailored medical or health advice without review by a qualified professional and without disclosing the use of artificial intelligence assistance and its potential limitations, as such actions may significantly impair the safety, well-being, or rights of others [52].

### Comprehensive and Professional Nature of the Supporting Materials for the Answers Provided by ChatGPT

The study found that ChatGPT-4o provided answers with higher completeness than Google Web Search and provided information drawn from professional information sources, such as peer-reviewed journals, open-access academic publications from IntechOpen, and clinical practice guidelines from Kidney Disease: Improving Global Outcomes. This finding is consistent with the findings of Oeding et al [53]. This may be because ChatGPT's training data primarily consists of a large volume of journal articles, books, reports, and other resources related to biomedical and health sciences, sourced from databases such as PubMed [54-56]. When responding to user queries, ChatGPT can quickly retrieve and integrate relevant information from its knowledge base. Based on this, compared with the previous study by Chen et al [57], we believe that ChatGPT can provide the general public with more professional answers regarding biologic therapy for SLE and help users better filter out false and misleading information from the internet.

### Aspects of Humanistic Care in ChatGPT's Answers

Based on content evaluation by 2 physicians, the study found that in ChatGPT-4o's responses to 20 questions about biologic therapy for SLE, 50% reflected aspects of humanistic care, such as medication risk warnings, expressions of concern for patients' physical and mental well-being, reminders about lifestyle adjustments, and considerations of medication affordability. This finding aligns with the research of Sorin et al [58] and Howe et al [59]. When providing answers regarding drug dosage and patient self-management of diseases, ChatGPT-4o includes recommendations at the end to seek guidance from health care providers. This reflects the rigor with which ChatGPT-4o delivers information to users. It also highlights the advantage of ChatGPT-4o, based on natural language processing technology, in effectively simulating human language and interaction patterns [60,61]. Compared with information on websites, which may contain logical inconsistencies or unclear statements, the way in which ChatGPT presents information is likely more accessible and comprehensible to the public. This helps to reduce the likelihood of patients being misled by vague or exaggerated information on advertising websites [57].

### Low Readability of ChatGPT's Answers

The study found that the texts generated by ChatGPT-4o regarding biologic therapy for SLE follow a general-specific-general structure, which aligns with human reading

habits. However, the content's readability was low. Based on the FRE and FKGL [27,28], the content provided by ChatGPT-4o regarding biologic therapy for SLE required a reading level equivalent to that of a college graduate [62], which is consistent with the findings of Ghanem et al [41]. The FRE score of Google Web Search was higher than that of ChatGPT-4o, which is consistent with the findings of Shen et al [63]. Possible reasons are as follows. First, ChatGPT was initially designed for a reading level comparable to that of high school students and above [29]. Second, ChatGPT's training data and cited sources include high-quality medical literature, ensuring the accuracy of the medical terminology used in its responses. However, the use of medical jargon may reduce the readability of the text [64]. Third, ChatGPT generates responses based on patterns learned from large-scale text data, which typically requires readers to have a reading proficiency above the sixth-grade level [29]. However, when given explicit instructions to respond at a sixth-grade reading level, ChatGPT's terminology density decreases, resulting in a significant improvement in the text's readability. Specifically, the Flesch Kincaid Grade Level score decreased from 11.7 to 10.5, while the FRE score increased substantially from 17.3 to 51.9, indicating a marked enhancement in readability. This finding is consistent with that of Shen et al [63] and Mohammadi et al [65]. This capability enables ChatGPT to provide responses tailored to different educational levels, which may have significant implications for improving health care information accessibility and reducing health care disparities [66].

## Implications

First, this study compares ChatGPT-4o and Google Web Search in providing information on biologic therapy for SLE, offering a reference for the future dissemination of SLE treatment knowledge based on LLMs.

Second, this study conducts an in-depth analysis of ChatGPT-4o's utility in providing information on biologic therapy for SLE, deepening understanding in both academia and industry regarding the potential of LLMs in delivering medical information and treatment recommendations.

Third, this study provides a reference for how users can use LLMs to access information on biologic therapy for SLE. It also offers insights into the effective use of LLMs for obtaining medical information to meet health care information needs.

## Acknowledgments

We appreciate the valuable comments provided by all reviewers. This study used the generative artificial intelligence tool of ChatGPT by OpenAI to generate common questions and answers related to biologic therapy for systemic lupus erythematosus. The method and process used to generate the responses are described in the relevant sections of the study.

## Data Availability

The datasets are available from the corresponding author upon reasonable request.

## Authors' Contributions

KL and BL are co-corresponding authors of this paper. Conceptualization was contributed by KL and BL. Methodology was contributed by KL and YP. Resources was handled by KL. Supervision was managed by KL and BL. Validation was contributed by KL, BL, and ZH. Investigation and data curation was handled by YP and LL. Formal analysis was contributed

## Limitations and Future Prospects

A few limitations should be noted in this study.

First, this study primarily focuses on common questions regarding biologic therapy for SLE, whereas in practice, users have a much broader range of concerns. Future studies can expand the scope of evaluated questions to gain a deeper understanding of ChatGPT's utility in providing information on biologic therapy for SLE.

Second, this study only compares the basic characteristics of questions provided by Google and ChatGPT, including question types, answer sources, the accuracy, completeness, and readability of answers. Future research could broaden the comparison dimensions by including more factors.

Third, this study lacks a comparative analysis of ChatGPT-4o's response characteristics and utility across different language contexts. Future research could focus on comparative studies across multiple languages.

Fourth, this study focuses on general internet users and does not account for differences across user groups. Future research could investigate the applicability and effectiveness of ChatGPT in meeting the information needs of various user groups.

Fifth, this study primarily focuses on evaluating information from a physician's perspective, without reflecting patient or user concerns and experiences. Future research could explore the effectiveness of ChatGPT in addressing the information needs of users or patients through surveys or interviews, providing a deeper understanding of how well ChatGPT-generated content meets these needs.

## Conclusion

By comparing the utility of ChatGPT-4o and Google Web Search in providing information on biologic therapy for SLE, this study finds that ChatGPT has the potential to be more effective in delivering accurate medical knowledge and supporting patient education. However, given the limitations of this study in terms of question scope, comparison dimensions, language types, and research perspectives, further in-depth comparative research is recommended.



by YP, LL, BL, and ZH. Visualization was contributed by LL. Writing—original draft was contributed by KL, YP, and LL. Writing—review and editing was contributed by KL, BL, YP, and LL.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Subtype distribution of 20 most common questions on biologic therapy for systemic lupus erythematosus from Google Web Search and ChatGPT-4o.

[PNG File (Portable Network Graphics File), 6 KB-Multimedia Appendix 1]

## Multimedia Appendix 2

The responses of ChatGPT-4o and Google Web Search.

[XLSX File (Microsoft Excel File), 52 KB-Multimedia Appendix 2]

## Multimedia Appendix 3

tableDistribution of website information sources for Google Web Search and ChatGPT-4o answers on biologic therapy for systemic lupus erythematosus.

[PNG File (Portable Network Graphics File), 193 KB-Multimedia Appendix 3]

## References

1. 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*. Jul 15, 2023;43(8):NP658-NP662. [doi: [10.1093/asj/sjad096](https://doi.org/10.1093/asj/sjad096)] [Medline: [37032521](https://pubmed.ncbi.nlm.nih.gov/37032521/)]
2. Hello GPT-4o. OpenAI. 2024. URL: <https://openai.com/index/hello-gpt-4o/> [Accessed 2024-06-13]
3. Choudhury A, Shamszare H. Investigating the impact of user trust on the adoption and use of ChatGPT: survey analysis. *J Med Internet Res*. Jun 14, 2023;25:e47184. [doi: [10.2196/47184](https://doi.org/10.2196/47184)] [Medline: [37314848](https://pubmed.ncbi.nlm.nih.gov/37314848/)]
4. Liu JL, Wang CY, Liu SR. Utility of ChatGPT in clinical practice. *J Med Internet Res*. Jun 2023;25:e48568. [doi: [10.2196/48568](https://doi.org/10.2196/48568)]
5. Ayoub NF, Lee YJ, Grimm D, Divi V. Head-to-head comparison of ChatGPT versus Google Search for medical knowledge acquisition. *Otolaryngol Head Neck Surg*. Jun 2024;170(6):1484-1491. [doi: [10.1002/ohn.465](https://doi.org/10.1002/ohn.465)] [Medline: [37529853](https://pubmed.ncbi.nlm.nih.gov/37529853/)]
6. Fernández-Pichel M, Pichel JC, Losada DE. Evaluating search engines and large language models for answering health questions. *NPJ Digit Med*. Mar 10, 2025;8(1):153. [doi: [10.1038/s41746-025-01546-w](https://doi.org/10.1038/s41746-025-01546-w)] [Medline: [40065094](https://pubmed.ncbi.nlm.nih.gov/40065094/)]
7. Ayık G, Ercan N, Demirtaş Y, Yıldırım T, Çakmak G. Evaluation of ChatGPT-4o's answers to questions about hip arthroscopy from the patient perspective. *Jt Dis Relat Surg*. Jan 2, 2025;36(1):193-199. [doi: [10.52312/jdrs.2025.1961](https://doi.org/10.52312/jdrs.2025.1961)] [Medline: [39719917](https://pubmed.ncbi.nlm.nih.gov/39719917/)]
8. Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*. Feb 2023;3(1):100105. [doi: [10.1016/j.tbench.2023.100105](https://doi.org/10.1016/j.tbench.2023.100105)]
9. Temsah MH, Jamal A, Alhasan K, et al. Transforming virtual healthcare: the potentials of ChatGPT-4omni in telemedicine. *Cureus*. May 2024;16(5):e61377. [doi: [10.7759/cureus.61377](https://doi.org/10.7759/cureus.61377)] [Medline: [38817799](https://pubmed.ncbi.nlm.nih.gov/38817799/)]
10. Ruksakulpiwat S, Kumar A, Ajibade A. Using ChatGPT in medical research: current status and future directions. *J Multidiscip Healthc*. 2023;16:1513-1520. [doi: [10.2147/JMDH.S413470](https://doi.org/10.2147/JMDH.S413470)] [Medline: [37274428](https://pubmed.ncbi.nlm.nih.gov/37274428/)]
11. Zhong N, Lu Z, Ge J, et al. *Internal Medicine*. People's Medical Publishing House; 2024.
12. Tian J, Zhang D, Yao X, Huang Y, Lu Q. Global epidemiology of systemic lupus erythematosus: a comprehensive systematic analysis and modelling study. *Ann Rheum Dis*. Mar 2023;82(3):351-356. [doi: [10.1136/ard-2022-223035](https://doi.org/10.1136/ard-2022-223035)] [Medline: [36241363](https://pubmed.ncbi.nlm.nih.gov/36241363/)]
13. Carter EE, Barr SG, Clarke AE. The global burden of SLE: prevalence, health disparities and socioeconomic impact. *Nat Rev Rheumatol*. Oct 2016;12(10):605-620. [doi: [10.1038/nrrheum.2016.137](https://doi.org/10.1038/nrrheum.2016.137)] [Medline: [27558659](https://pubmed.ncbi.nlm.nih.gov/27558659/)]
14. Insights SG. In a Push to Limit Use of Glucocorticoids, US and EU5 Rheumatologists are Increasingly Turning to GSK's Benlysta and AstraZeneca's Saphnelo for the Treatment of Moderate-to-Severe Systemic Lupus Erythematosus, According to Spherix Global In. URL: <https://www.globenewswire.com/news-release/2023/10/04/2754325/0/en/In-a-Push-to-Limit-Use-of-Glucocorticoids-US-and-EU5-Rheumatologists-are-Increasingly-Turning-to-GSK-s-Benlysta-and-AstraZeneca-s-Saphnelo-for-the-Treatment-of-Moderate-to-Severe-S.html> [Accessed 2023-06-15]
15. Kirou KA, Dall Era M, Aranow C, Anders HJ. Belimumab or anifrolumab for systemic lupus erythematosus? A risk-benefit assessment. *Front Immunol*. 2022;13:980079. [doi: [10.3389/fimmu.2022.980079](https://doi.org/10.3389/fimmu.2022.980079)] [Medline: [36119023](https://pubmed.ncbi.nlm.nih.gov/36119023/)]

16. Geng Y, Wu L, Xie Q, et al. 2024 Chinese expert consensus on the use of biologic therapy for systemic lupus erythematosus. *Chinese Journal of Rheumatology*. 2024;28(2):78-92.
17. Liu X, Song Y, Wan L, Du R. Knowledge, attitudes, and practices among patients with systemic lupus erythematosus toward disease management and biologic therapy. *J Multidiscip Healthc*. 2024;17:937-947. [doi: [10.2147/JMDH.S444619](https://doi.org/10.2147/JMDH.S444619)] [Medline: [38455274](https://pubmed.ncbi.nlm.nih.gov/38455274/)]
18. Waldron N, Brown S, Hewlett S, Elliott B, McHugh N, McCabe C. “It’s more scary not to know”: a qualitative study exploring the information needs of patients with systemic lupus erythematosus at the time of diagnosis. *Musculoskeletal Care*. Dec 2011;9(4):228-238. [doi: [10.1002/msc.221](https://doi.org/10.1002/msc.221)] [Medline: [21993962](https://pubmed.ncbi.nlm.nih.gov/21993962/)]
19. Ichikawa T, Kishida D, Shimojima Y, et al. Impact of online health information-seeking behavior on shared decision-making in patients with systemic lupus erythematosus: The TRUMP<sup>2</sup>-SLE project. *Lupus (Los Angel)*. Oct 2023;32(11):1258-1266. [doi: [10.1177/09612033231200104](https://doi.org/10.1177/09612033231200104)] [Medline: [37651318](https://pubmed.ncbi.nlm.nih.gov/37651318/)]
20. Majority of Adults Look Online for Health Information. Pew Research Center. 2013. URL: <https://www.pewresearch.org/short-reads/2013/02/01/majority-of-adults-look-online-for-health-information/> [Accessed 2024-06-15]
21. Jacquemart P, Zweigenbaum P. Towards a medical question-answering system: a feasibility study. *Stud Health Technol Inform*. 2003;95:463-468. [Medline: [14664030](https://pubmed.ncbi.nlm.nih.gov/14664030/)]
22. Teixeira Lopes C, Ribeiro C. Comparative evaluation of web search engines in health information retrieval. *Online Information Review*. Nov 29, 2011;35(6):869-892. [doi: [10.1108/14684521111193175](https://doi.org/10.1108/14684521111193175)]
23. Huang Y, Huang JX. Exploring ChatGPT for next-generation information retrieval: opportunities and challenges. *WEB*. 2024;22(1):31-44. [doi: [10.3233/WEB-230363](https://doi.org/10.3233/WEB-230363)]
24. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatiopoulos V, Kaklamanos EG. Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: comparative mixed methods study. *J Med Internet Res*. Dec 28, 2023;25:e51580. [doi: [10.2196/51580](https://doi.org/10.2196/51580)] [Medline: [38009003](https://pubmed.ncbi.nlm.nih.gov/38009003/)]
25. HopkinsAM, LoganJM, KichenadasseG, Sorich MJ. Artificial intelligence chatbots will revolutionize how cancer patients access information: ChatGPT represents a paradigm-shift. *JNCI Cancer Spectr*. Aug 31, 2023;7(2):pkad010. [doi: [10.1093/jncics/pkad010](https://doi.org/10.1093/jncics/pkad010)] [Medline: [36808255](https://pubmed.ncbi.nlm.nih.gov/36808255/)]
26. Eltorai AEM, Han A, Truntzer J, Daniels AH. Readability of patient education materials on the American Orthopaedic Society for Sports Medicine website. *Phys Sportsmed*. Nov 2014;42(4):125-130. [doi: [10.3810/psm.2014.11.2099](https://doi.org/10.3810/psm.2014.11.2099)] [Medline: [25419896](https://pubmed.ncbi.nlm.nih.gov/25419896/)]
27. Daraz L, Morrow AS, Ponce OJ, et al. Readability of online health information: a meta-narrative systematic review. *Am J Med Qual*. 2018;33(5):487-492. [doi: [10.1177/1062860617751639](https://doi.org/10.1177/1062860617751639)] [Medline: [29345143](https://pubmed.ncbi.nlm.nih.gov/29345143/)]
28. Massie PL, Arshad SA, Auyang ED. Readability of American Society of Metabolic Surgery’s patient information publications. *J Surg Res*. Jan 2024;293:727-732. [doi: [10.1016/j.jss.2023.09.018](https://doi.org/10.1016/j.jss.2023.09.018)] [Medline: [37862852](https://pubmed.ncbi.nlm.nih.gov/37862852/)]
29. Moons P, Van Bulck L. Using ChatGPT and Google Bard to improve the readability of written patient information: a proof of concept. *Eur J Cardiovasc Nurs*. Mar 12, 2024;23(2):122-126. [doi: [10.1093/eurjcn/zvad087](https://doi.org/10.1093/eurjcn/zvad087)] [Medline: [37603843](https://pubmed.ncbi.nlm.nih.gov/37603843/)]
30. Huang C, Hong D, Chen L, Chen X. Assess the precision of ChatGPT’s responses regarding systemic lupus erythematosus (SLE) inquiries. *Skin Res Technol*. Oct 2023;29(10):e13500. [doi: [10.1111/srt.13500](https://doi.org/10.1111/srt.13500)] [Medline: [37881050](https://pubmed.ncbi.nlm.nih.gov/37881050/)]
31. How can i access GPT-4, GPT-4 turbo, GPT-4o, and GPT-4o mini? OpenAI. 2024. URL: [https://help.openai.com/en/articles/7102672-how-can-i-access-gpt-4-gpt-4-turbo-gpt-4o-and-gpt-4o-mini?utm\\_medium=email&utm\\_source=transaction](https://help.openai.com/en/articles/7102672-how-can-i-access-gpt-4-gpt-4-turbo-gpt-4o-and-gpt-4o-mini?utm_medium=email&utm_source=transaction) [Accessed 2024-06-15]
32. O’Mathúna DP. How should clinicians engage with online health information? *AMA J Ethics*. Nov 1, 2018;20(11):E1059-1066. [doi: [10.1001/amajethics.2018.1059](https://doi.org/10.1001/amajethics.2018.1059)] [Medline: [30499435](https://pubmed.ncbi.nlm.nih.gov/30499435/)]
33. Dubin JA, Bains SS, Chen Z, et al. Using a Google Web Search analysis to assess the utility of ChatGPT in total joint arthroplasty. *J Arthroplasty*. Jul 2023;38(7):1195-1202. [doi: [10.1016/j.arth.2023.04.007](https://doi.org/10.1016/j.arth.2023.04.007)] [Medline: [37040823](https://pubmed.ncbi.nlm.nih.gov/37040823/)]
34. Chelli M, Descamps J, Lavoué V, et al. Hallucination rates and reference accuracy of ChatGPT and Bard for systematic reviews: comparative analysis. *J Med Internet Res*. May 22, 2024;26:e53164. [doi: [10.2196/53164](https://doi.org/10.2196/53164)] [Medline: [38776130](https://pubmed.ncbi.nlm.nih.gov/38776130/)]
35. WMA declaration of helsinki - ethical principles for medical research involving human participants. World Medical Association. 2025. URL: <https://www.wma.net/policies-post/wma-declaration-of-helsinki/> [Accessed 2025-04-11]
36. Rothwell JD. *Mixed Company: Communication in Small Groups*. Wadsworth Cengage Learning; 2013. ISBN: 1111346852
37. López-Jornet P, Camacho-Alonso F. The quality of internet sites providing information relating to oral cancer. *Oral Oncol*. Sep 2009;45(9):e95-8. [doi: [10.1016/j.oraloncology.2009.03.017](https://doi.org/10.1016/j.oraloncology.2009.03.017)] [Medline: [19457707](https://pubmed.ncbi.nlm.nih.gov/19457707/)]

38. Hahn BH, McMahon MA, Wilkinson A, et al. American College of Rheumatology guidelines for screening, treatment, and management of lupus nephritis. *Arthritis Care Res (Hoboken)*. Jun 2012;64(6):797-808. [doi: [10.1002/acr.21664](https://doi.org/10.1002/acr.21664)]
39. Goodman RS, Patrinely JR, Stone CA Jr, et al. Accuracy and reliability of chatbot responses to physician questions. *JAMA Netw Open*. Oct 2, 2023;6(10):e2336483. [doi: [10.1001/jamanetworkopen.2023.36483](https://doi.org/10.1001/jamanetworkopen.2023.36483)] [Medline: [37782499](https://pubmed.ncbi.nlm.nih.gov/37782499/)]
40. Morris LA, Myers A, Thilman DG. Application of the readability concept to patient-oriented drug information. *Am J Hosp Pharm*. Nov 1980;37(11):1504-1509. [Medline: [7211852](https://pubmed.ncbi.nlm.nih.gov/7211852/)]
41. Ghanem YK, Rouhi AD, Al-Houssan A, et al. Dr. Google to Dr. ChatGPT: assessing the content and quality of artificial intelligence-generated medical information on appendicitis. *Surg Endosc*. May 2024;38(5):2887-2893. [doi: [10.1007/s00464-024-10739-5](https://doi.org/10.1007/s00464-024-10739-5)] [Medline: [38443499](https://pubmed.ncbi.nlm.nih.gov/38443499/)]
42. Jindal P, MacDermid JC. Assessing reading levels of health information: uses and limitations of Flesch formula. *Educ Health (Abingdon)*. 2017;30(1):84-88. [doi: [10.4103/1357-6283.210517](https://doi.org/10.4103/1357-6283.210517)] [Medline: [28707643](https://pubmed.ncbi.nlm.nih.gov/28707643/)]
43. Solnyshkina M, Zamaletdinov R, Gorodetskaya L, Gabitov A. Evaluating text complexity and Flesch-Kincaid grade level. *J Soc Stud Educ Res*. 2017;8(3):238-248. URL: <https://files.eric.ed.gov/fulltext/EJ1162266.pdf>
44. Cohen SA, Brant A, Fisher AC, Pershing S, Do D, Pan C. Dr. Google vs. Dr. ChatGPT: exploring the use of artificial intelligence in ophthalmology by comparing the accuracy, safety, and readability of responses to frequently asked patient questions regarding cataracts and cataract surgery. *Semin Ophthalmol*. Aug 2024;39(6):472-479. [doi: [10.1080/08820538.2024.2326058](https://doi.org/10.1080/08820538.2024.2326058)] [Medline: [38516983](https://pubmed.ncbi.nlm.nih.gov/38516983/)]
45. Srinivasan D. Why Google dominates advertising markets. *Stanf Technol Law Rev*. 2020;24(1):55. URL: <https://law.stanford.edu/wp-content/uploads/2020/12/Srinivasan-FINAL-Why-Google-Dominates-Advertising-Markets.pdf> [Accessed 2025-08-18]
46. Bojazar R, Do TP, Hansen JM, Dodick DW, Ashina M. Googling migraine: a study of Google as an information resource of migraine management. *Cephalalgia*. Dec 2020;40(14):1633-1644. [doi: [10.1177/0333102420942241](https://doi.org/10.1177/0333102420942241)] [Medline: [32703015](https://pubmed.ncbi.nlm.nih.gov/32703015/)]
47. Zenone M, Schalkwyk M, Maani N, et al. Google Allows Snake-Oil Advertisers to Target Cancer Patient Queries. *Brewminate: A Bold Blend of News and Ideas*. 2025. URL: <https://brewminate.com/google-allows-snake-oil-advertisers-to-target-cancer-patient-queries/> [Accessed 2025-06-07]
48. Pleasants E, Guendelman S, Weidert K, Prata N. Quality of top webpages providing abortion pill information for Google searches in the USA: an evidence-based webpage quality assessment. *PLoS ONE*. 2021;16(1):e0240664. [doi: [10.1371/journal.pone.0240664](https://doi.org/10.1371/journal.pone.0240664)] [Medline: [33476340](https://pubmed.ncbi.nlm.nih.gov/33476340/)]
49. Wijaya D, Daniawan B, Gunawan Y. Search engine optimization (SEO) as a promotional media on Google Search. *bit-Tech*. 2021;4(1):31-39. URL: <https://jurnal.kdi.or.id/index.php/bt/issue/view/24> [Accessed 2025-08-18] [doi: [10.32877/bt.v4i1.237](https://doi.org/10.32877/bt.v4i1.237)]
50. Lewandowski D, Sünkler S, Yagci N. The influence of search engine optimization on google's results: a multi-dimensional approach for detecting SEO. Presented at: WebSci '21: Proceedings of the 13th ACM Web Science Conference 2021; Jun 21-25, 2021; Virtual Event United Kingdom. 2021.[doi: [10.1145/3447535.3462479](https://doi.org/10.1145/3447535.3462479)]
51. Kasper D, Fauci A, Hauser S, Longo D, Jameson J, Loscalzo J. *Harrison's Principles of Internal Medicine*, 19e. McGraw-hill New York, NY, USA; 2015.
52. Usage policies. OpenAI. URL: <https://openai.com/policies/usage-policies/> [Accessed 2025-01-07]
53. Oeding JF, Lu AZ, Mazzucco M, et al. ChatGPT-4 performs clinical information retrieval tasks using consistently more trustworthy resources than does Google Search for queries concerning the latarjet procedure. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*. Mar 2025;41(3):588-597. [doi: [10.1016/j.arthro.2024.05.025](https://doi.org/10.1016/j.arthro.2024.05.025)]
54. PubMed. National Institutes of Health. 2023. URL: <https://pubmed.ncbi.nlm.nih.gov/about/> [Accessed 2025-08-16]
55. Tian S, Jin Q, Yeganova L, et al. Opportunities and challenges for ChatGPT and large language models in biomedicine and health. *Brief Bioinformatics*. Nov 22, 2023;25(1). [doi: [10.1093/bib/bbad493](https://doi.org/10.1093/bib/bbad493)]
56. AlSamhori ARF, AlSamhori JF, AlSamhori AF. ChatGPT role in a medical survey. *HYMR*. 2023;1(2). [doi: [10.59707/hymrTFFP5435](https://doi.org/10.59707/hymrTFFP5435)]
57. Chen L, Li H, Su Y, et al. Using a Google Web Search analysis to assess the utility of ChatGPT in stem cell therapy. *Stem Cells Transl Med*. Jan 12, 2024;13(1):60-68. [doi: [10.1093/stcltm/szad074](https://doi.org/10.1093/stcltm/szad074)] [Medline: [37936506](https://pubmed.ncbi.nlm.nih.gov/37936506/)]
58. Sorin V, Brin D, Barash Y, et al. Large language models and empathy: systematic review. *J Med Internet Res*. Dec 11, 2024;26:e52597. [doi: [10.2196/52597](https://doi.org/10.2196/52597)] [Medline: [39661968](https://pubmed.ncbi.nlm.nih.gov/39661968/)]
59. Howe PDL, Fay N, Saletta M, Hovy E. ChatGPT's advice is perceived as better than that of professional advice columnists. *Front Psychol*. 2023;14:1281255. [doi: [10.3389/fpsyg.2023.1281255](https://doi.org/10.3389/fpsyg.2023.1281255)] [Medline: [38078232](https://pubmed.ncbi.nlm.nih.gov/38078232/)]
60. Praveen SV, Vajrobol V. Can ChatGPT be trusted for consulting? Uncovering doctor's perceptions using deep learning techniques. *Ann Biomed Eng*. Oct 2023;51(10):2116-2119. [doi: [10.1007/s10439-023-03245-7](https://doi.org/10.1007/s10439-023-03245-7)] [Medline: [37208451](https://pubmed.ncbi.nlm.nih.gov/37208451/)]

61. Shahriar S, Lund BD, Mannuru NR, et al. Putting GPT-4o to the sword: a comprehensive evaluation of language, vision, speech, and multimodal proficiency. *Appl Sci (Basel)*. 2024;14(17):7782. [doi: [10.3390/app14177782](https://doi.org/10.3390/app14177782)]
62. Doğan L, Özçakmakçı GB, Yılmaz İE. The performance of chatbots and the AAPOS website as a tool for amblyopia education. *J Pediatr Ophthalmol Strabismus*. 2024;61(5):325-331. [doi: [10.3928/01913913-20240409-01](https://doi.org/10.3928/01913913-20240409-01)] [Medline: [38661309](https://pubmed.ncbi.nlm.nih.gov/38661309/)]
63. Shen SA, Perez-Heydrich CA, Xie DX, Nellis JC. ChatGPT vs. web search for patient questions: what does ChatGPT do better? *Eur Arch Otorhinolaryngol*. Jun 2024;281(6):3219-3225. [doi: [10.1007/s00405-024-08524-0](https://doi.org/10.1007/s00405-024-08524-0)] [Medline: [38416195](https://pubmed.ncbi.nlm.nih.gov/38416195/)]
64. Lang IA, King A, Boddy K, et al. Jargon and readability in plain language summaries of health research: cross-sectional observational study. *J Med Internet Res*. Jan 13, 2025;27:e50862. [doi: [10.2196/50862](https://doi.org/10.2196/50862)] [Medline: [39805102](https://pubmed.ncbi.nlm.nih.gov/39805102/)]
65. Mohammadi SS, Khatri A, Jain T, et al. Evaluation of the appropriateness and readability of ChatGPT-4 responses to patient queries on uveitis. *Ophthalmol Sci*. 2025;5(1):100594. [doi: [10.1016/j.xops.2024.100594](https://doi.org/10.1016/j.xops.2024.100594)] [Medline: [39435137](https://pubmed.ncbi.nlm.nih.gov/39435137/)]
66. Van Bulck L, Moons P. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. *Eur J Cardiovasc Nurs*. Jan 12, 2024;23(1):95-98. [doi: [10.1093/eurjcn/zvad038](https://doi.org/10.1093/eurjcn/zvad038)] [Medline: [37094282](https://pubmed.ncbi.nlm.nih.gov/37094282/)]

## Abbreviations

**LLM:** large language model

**SLE:** systemic lupus erythematosus

**FRE:** Flesch Reading Ease

**FKGL:** Flesch-Kincaid Grade Level

*Edited by Amaryllis Mavragani; peer-reviewed by Metwaly Eldakar, ZhengPing Huang; submitted 24.04.2025; final revised version received 18.06.2025; accepted 03.07.2025; published 28.08.2025*

*Please cite as:*

*Li K, Peng Y, Li L, Liu B, Huang Z*

*Evaluating ChatGPT's Utility in Biologic Therapy for Systemic Lupus Erythematosus: Comparative Study of ChatGPT and Google Web Search*

*JMIR Form Res 2025;9:e76458*

*URL: <https://formative.jmir.org/2025/1/e76458>*

*doi: [10.2196/76458](https://doi.org/10.2196/76458)*

© Kai Li, Yunfei Peng, Luyi Li, Bo Liu, Zhijian Huang. Originally published in JMIR Formative Research (<https://formative.jmir.org>), 28.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Formative Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://formative.jmir.org>, as well as this copyright and license information must be included.