

Evaluating Artificial Intelligence (AI)-Generated Patient Education Guides on Epilepsy: A Cross-Sectional Study of ChatGPT and Google Gemini

Review began 10/10/2024
Review ended 10/29/2024
Published 11/07/2024

© Copyright 2024

Arora et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.73212

Shiv Arora ¹, Meghna Ramesh ², Aye Thandar Moe ³, Tapan Giri ⁴, Kaksha Parrikh ⁵, Hima Varsha Challa ^{6, 7}

1. General Surgery, Sardar Patel Medical College, Bikaner, IND 2. Internal Medicine, Asian Institute of Gastroenterology Hospitals, Hyderabad, IND 3. Internal Medicine, University of Medicine 1 Yangon, Myanmar, MMR 4. Internal Medicine, Byramjee Jeejeebhoy Government Medical College, Pune, IND 5. Medicine, Government Medical College, Bhavnagar, Bhavnagar, IND 6. Internal Medicine, Sri Ramachandra Institute of Higher Education and Research, Chennai, IND 7. Internal Medicine, University Hospitals of Leicester NHS Trust, Leicester, GBR

Corresponding author: Hima Varsha Challa, challahima68@gmail.com

Abstract

Introduction

Epilepsy is a chronic disorder that requires patient education for management and to avoid triggers and complications. This study aims to evaluate and compare the effectiveness of two artificial intelligence (AI) tools, ChatGPT (version 3.5, OpenAI, Inc., San Francisco, United States) and Google Gemini (version 1.5, Google LLC, Mountain View, California, United States), in generating patient education guides for epilepsy disorders.

Methodology

A patient education guide was generated on ChatGPT and Google Gemini. The study analyzed the sentence count, readability, and ease of understanding using the Flesch-Kincaid calculator, examined similarity using the QuillBot plagiarism tool, and assessed reliability using a modified DISCERN score. Statistical analysis included an unpaired T-test where a P-value <0.05 is considered significant.

Results

There was no statistically significant difference between ChatGPT and Google Gemini in terms of word count (p=0.75), sentence count (p=0.96), average words per sentence (p=0.66), grade level (p=0.67), similarity% (p=0.57), and reliability scores (p=0.42). Ease scores generated by ChatGPT and Google Gemini were 38.6 and 43.6 for generalized tonic-clonic seizures (GTCS), 18.7 and 45.5 for myoclonic seizures, and 22.4 and 55.8 for status epilepticus, respectively, showing Google Gemini generated responses notably better (p=0.0493). The average syllables per word (p=0.035) were appreciably lower for Google Gemini-generated responses, with 1.8 for GTCS and myoclonic, 1.7 for status epilepticus against 1.9 for GTCS, 2 for myoclonic, and 2.1 for status epilepticus for ChatGPT responses.

Conclusions

A significant difference was seen in only two parameters. Further improvement in AI tools is necessary to provide effective guides.

Categories: Neurology, Internal Medicine, Medical Education

Keywords: artificial intelligence, chatgpt, education guide, epilepsy, generalized tonic-clonic seizures (gtcs), google gemini, myoclonic seizures, seizures, status epilepticus

Introduction

Epilepsy, being one of the most common serious brain conditions, is characterized by the recurrence of unprovoked seizures [1]. It affects over 70 million people worldwide [2]. Patient education in specific seizure types like generalized tonic-clonic seizures (GTCS), myoclonic seizures, and status epilepticus is important as it helps in the early identification of complications and maintenance of medications, thereby preventing complications/early mortality or morbidity. It can also help in the faster resolution of these diseases.

Artificial intelligence (AI) tools have emerged as potential game-changers in patient education, offering readily accessible information on a wide range of diseases. AI also has the potential to revolutionize personalized medicine and enhance population health management. It can also provide virtual health assistants, support mental health care, improve patient education, and influence patient-physician trust [3]. The challenge lies in verifying the accuracy of these tools. There is also a lack of personal connection to doctors. Additionally, there can be an overreliance on these tools for diagnosis and treatment. ChatGPT (version 3.5, OpenAI, Inc., San Francisco, United States) is a revolutionary technology that uses advanced AI

How to cite this article

Arora S, Ramesh M, Moe A, et al. (November 07, 2024) Evaluating Artificial Intelligence (AI)-Generated Patient Education Guides on Epilepsy: A Cross-Sectional Study of ChatGPT and Google Gemini. Cureus 16(11): e73212. DOI 10.7759/cureus.73212

techniques to generate natural language responses to a given prompt or input. It has been used across various fields, from natural language processing to customer service to content creation [4]. Google Gemini (version 1.5, Google LLC, Mountain View, California, United States) is a powerful AI model from Google that can understand text, images, videos, and audio. Google Gemini showcases the latest advancements in AI and has attracted attention for its capabilities, which are comparable to the widely popular ChatGPT [5].

The goal is to assess the potential of ChatGPT and Google Gemini to improve the quality and accessibility of patient counseling for individuals living with specific seizure types such as GTCS, myoclonic seizures, and status epilepticus.

Aims and objectives

To evaluate the effectiveness of ChatGPT and Google Gemini in generating patient education guides for GTCS, myoclonic seizures, and status epilepticus, and to assess the accuracy, comprehensiveness, and readability of AI-generated materials. This manuscript also aims to determine the potential clinical utility of these AI tools in epilepsy patient education and management.

Materials And Methods

A cross-sectional original research study was conducted over one week in June 2024. Since no human participants were involved in this study, ethics committee approval was not needed.

Firstly, the responses were generated. Three common diseases in neurology were selected, namely GTCS, myoclonic seizures, and status epilepticus. Two AI tools were selected, namely ChatGPT and Google Gemini, for the generation of brochures for patient education. ChatGPT version 3.5 and Google Gemini version 1.5 were used on June 11th, 2024 [6]. The following prompt was given to both AI tools: "Write a patient education guide for [disease name]." Responses generated were collected in a Microsoft Word (Microsoft Corp., Redmond, WA, US) document and were graded using various tools: the Flesch-Kincaid calculator was used for word count, sentence count, ease of understanding and readability of the information generated, similarity of the content was checked using the QuillBot Plagiarism Tool, and reliability of scientific text using a modified DISCERN score [7,8,9]. The modified DISCERN score is a tool to measure reliability and accuracy with a total score of five points. The higher scores indicate higher reliability [9].

Statistical analysis data was exported to a Microsoft Excel sheet. Statistical analysis was done using R version 4.3.2. R core team (2023) (R: a language and environment for statistical computing; R Foundation for Statistical Computing, Vienna, Austria, <https://www.R-project.org/>). The responses generated by ChatGPT and Google Gemini were compared using an unpaired T-test. A P-value <0.05 was considered significant. The correlation between ease score and reliability score was compared using Pearson's coefficient of correlation.

Results

ChatGPT and Google Gemini were used to generate brochures on patient education for GTCS, myoclonic seizures, and status epilepticus.

Table 1 shows that there was no significant difference in the word count ($p=0.753$), sentence count ($p=0.9626$), average words per sentence ($p=0.6633$), grade level ($p=0.6714$), similarity% ($p=0.5722$), and reliability scores ($p=0.4226$) between ChatGPT and Google Gemini.

Variables	ChatGPT		Google Gemini		P-value ⁺
	Mean	Standard Deviation	Mean	Standard Deviation	
Words	392.30	128.75	419.70	40.50	0.7543
Sentences	47.00	32.19	48.00	7.81	0.9626
Average words per sentence	10.93	6.74	8.93	1.85	0.6633
Average syllables per word	2.00	0.10	1.77	0.06	0.0357*
Grade level	12.27	2.83	10.90	4.28	0.6714
Ease score	26.57	10.58	48.30	6.56	0.0493*
Similarity %	37.93	34.70	23.80	16.79	0.5722
Reliability score	2.67	0.58	3.00	0.00	0.4226

TABLE 1: Characteristics of responses generated by ChatGPT and Google Gemini

+ t-test. P-values <0.05 are considered statistically significant.

However, there was a significant difference in the average syllables per word (p=0.0357) and the ease score (p=0.0493) between ChatGPT and Google Gemini. The ease score was significantly higher for Google Gemini-generated responses as compared to ChatGPT, while the average syllables per word were significantly less for Google Gemini-generated responses as compared to ChatGPT.

Figure 1 shows that the Flesch-Kincaid grade level of the response generated by Google Gemini was higher (9.9) than ChatGPT (9.7) for the topic of generalized tonic-clonic seizures. Similarly, the Flesch-Kincaid grade level of the response generated by Google Gemini was higher (15.59) as compared to ChatGPT (15.3) for the topic of myoclonic seizures. However, for the topic of status epilepticus, the Flesch-Kincaid grade level of the response generated by ChatGPT (11.8) was higher than Google Gemini (7.2).

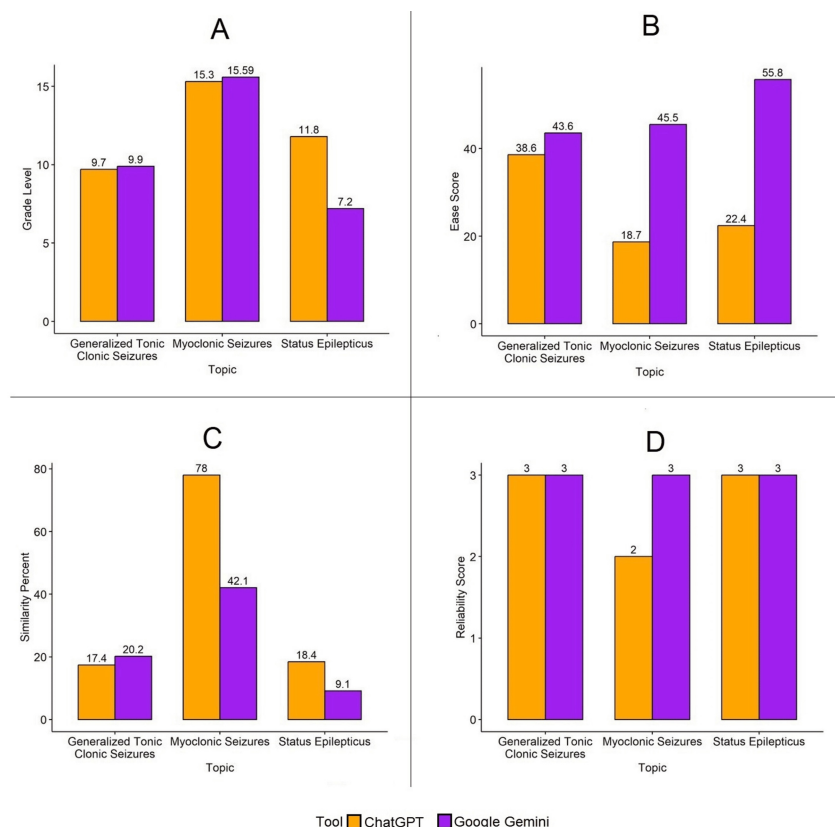


FIGURE 1: Graphical representation of comparison between grade level, ease score, similarity percent, and reliability score, for the patient education guide generated by ChatGPT and Google Gemini

a: comparison of grade level; b: comparison of ease score; c: comparison of similarity%; d: comparison of reliability score

The Flesch reading ease score of the patient education brochure generated by Google Gemini on the topic of GTCS was higher (43.8) as compared to ChatGPT (38.6), with both responses having a college reading level (difficult to read). For the topic of myoclonic seizures, the Flesch reading ease score of the response generated by Google Gemini was higher (45.5), corresponding to a college reading level (difficult to read) as compared to the response generated by ChatGPT (18.7), which had a college graduate reading level (very difficult to read).

For the topic of status epilepticus, the patient education brochure generated by Google Gemini had a higher Flesch reading ease score (55.8) with a corresponding reading level of high school (difficult to read) as compared to the response generated by ChatGPT (22.4), which had a college graduate reading level (very difficult to read).

The similarity percentage of the response generated by Google Gemini was higher (20.2%) as compared to ChatGPT (17.4%) for the topic of GTCS. Conversely, for the topic of myoclonic seizures, the similarity percentage of the patient education guide generated by ChatGPT was higher (78%) as compared to Google Gemini (42.1%). Likewise, the similarity percentage of the response generated by ChatGPT was higher (18.4%) as compared to Google Gemini (9.1%) for the topic of status epilepticus.

The modified DISCERN score (reliability score) was the same (three points) for the patient education guides generated by both ChatGPT and Google Gemini on the topics of GTCS and status epilepticus. However, the reliability score was higher for the response generated by Google Gemini (three points) as compared to ChatGPT (two points) for the topic of myoclonic seizures.

Discussion

A cross-sectional study conducted to compare responses generated by two AI tools, ChatGPT and Google Gemini, for brochures on patient education for GTCS, myoclonic seizures, and status epilepticus revealed that there was no significant difference in the word count, sentence count, average words per sentence, grade level, similarity%, and reliability scores between ChatGPT and Google Gemini. Ease score was higher

in Google Gemini as compared to ChatGPT.

AI can play a significant role in a patient's education by providing personalized, accessible, and accurate health information. Its ability to simplify health concerns can make it a powerful tool in the education of patients. Short and concise texts are always simple to read and retain. For easy-to-read brochures, the level of understanding (ease score) should be such that even a high school person can understand. In this study, both ChatGPT and Google Gemini generated responses at the college level of understanding for GTCS. For both myoclonic seizures and status epilepticus, ChatGPT is at a very difficult level, and Google Gemini has college-level difficulty.

This finding can be compared to a study that used an AI dialogue platform to improve the readability of orthopedic surgery-related education materials to the fifth-grade level and found that it can improve the reading accessibility to orthopedic surgery-related online study material [10]. Another study took 71 articles from patient education materials published by the American Academy of Otolaryngology, and ChatGPT4 was successfully used to translate those materials to sixth-grade reading level [11]. Twenty-one patient education materials on aortic stenosis were successfully used by another study to make it readable at the 10th-12th reading level, highlighting ChatGPT's role in the future for cardiac health literacy [12]. Another study done in 2016 found that the readability of patient education material from the American Academy of Orthopedic Surgeons (AAOS) has improved from 2008-2016 [13]. A study found AI has revolutionized patient education in dentistry by identifying articles from 2018 to 2024, encompassing empirical evidence and conceptual frameworks related to AI in dental patient engagement and communication [14]. In one study, ChatGPT, Microsoft Copilot (Microsoft Corp., Redmond, WA, US), Google Gemini, and Meta AI (Meta Platforms, Inc., Menlo Park, California, United States) were asked 10 questions regarding cardiac catheterization that were developed by website-based patient education materials. The responses generated had an overall mean reading between the 11th and 13th grade levels of education materials [15]. Thus, both ChatGPT and Google Gemini need to be updated to improve the difficulty level in patient education guides for GTCS, status epilepticus, and myoclonic seizures.

Similarity percentage relates to copied content from pre-existing literature, and this is used to avoid plagiarism. Plagiarism is detrimental in medical publications as it can be related to the credibility of the author and faith in medical updates. With an increase in medical publications, all need to follow publication ethics [16]. The similarity% of the patient education guide for GTCS was higher in Google Gemini (20.2%) as compared to ChatGPT (17.4%), while ChatGPT has a higher similarity% (78 and 18.4%) than Google Gemini (42.1 and 9.1%) in both myoclonic seizures and status epilepticus, respectively, with the expected similarity limit being 15%, which is only under the limit in the Google Gemini guide for status epilepticus. Patient education articles generated by ChatGPT in common dermatologic conditions have a similarity of 27%, which is also higher than the expected limit [17].

A modified DISCERN score is used for reliability of online/media content which comprises five questions, with each affirmative response being assigned a score of one. In this study, both ChatGPT and Google Gemini received a score of three for content related to GTCS and status epilepticus. For myoclonus seizures, ChatGPT received a score of two, while Google Gemini received a score of three. One similar study on patient education on sports surgery-related diseases by using ChatGPT-4 found that it has the potential to provide quality information with a DISCERN score of 44.75 points [18].

Limitations

This study used only two AI tools. In the future, we need to assess more tools and study more diseases. ChatGPT 3.5 is an older version of artificial intelligence, and it may not provide updated content. Although these educational guides aim to bridge the gap, medical professionals possess the necessary training to ensure patient-centered management, a skill that these AI models lack.

Conclusions

This study highlights that there is no significant difference in the average ease score, grade score, and reliability score of responses generated by the two AI tools for patient information brochures on GTCS, myoclonic seizures, and status epilepticus. The ease score was higher in Google Gemini between the two software. Similarity% is also higher than the limit of 15% in all education guides except for Google Gemini for status epilepticus.

More advancements or updated AI tools are required to enhance the readability of these neurological conditions.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Hima Varsha Challa, Shiv Arora, Meghna Ramesh, Aye Thandar Moe, Tapan Giri, Kaksha Parrikh

Acquisition, analysis, or interpretation of data: Hima Varsha Challa, Shiv Arora, Meghna Ramesh, Aye Thandar Moe, Tapan Giri, Kaksha Parrikh

Drafting of the manuscript: Hima Varsha Challa, Shiv Arora, Meghna Ramesh, Aye Thandar Moe, Tapan Giri, Kaksha Parrikh

Critical review of the manuscript for important intellectual content: Hima Varsha Challa, Shiv Arora, Meghna Ramesh, Aye Thandar Moe, Tapan Giri, Kaksha Parrikh

Supervision: Hima Varsha Challa, Shiv Arora, Meghna Ramesh, Aye Thandar Moe, Tapan Giri, Kaksha Parrikh

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue.

Animal subjects: All authors have confirmed that this study did not involve animal subjects or tissue.

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Beghi E, Giussani G, Sander JW: The natural history and prognosis of epilepsy . *Epileptic Disord*. 2015, 17:243-53. [10.1684/epd.2015.0751](https://doi.org/10.1684/epd.2015.0751)
2. Thijs RD, Surges R, O'Brien TJ, Sander JW: Epilepsy in adults. *Lancet*. 2019, 393:689-701.
3. Alowais SA, Alghamdi SS, Alsuhbany N, et al.: Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Med Educ*. 2023, 23:689. [10.1186/s12909-023-04698-z](https://doi.org/10.1186/s12909-023-04698-z)
4. Kalla D, Smith N, Samaah F, Kuraku S: Study and analysis of Chat GPT and its impact on different fields of study. *Social Science Research Network*. 2023, 8:3.
5. Lee TJ, Campbell DJ, Patel S, Hossain A, Radfar N, Siddiqui E, Gardin JM: Unlocking health literacy: the ultimate guide to hypertension education from ChatGPT versus Google Gemini. *Cureus*. 2024, 16:e59898. [10.7759/cureus.59898](https://doi.org/10.7759/cureus.59898)
6. Shukla R, Mishra AK, Banerjee N, Verma A: The comparison of ChatGPT 3.5, Microsoft Bing, and Google Gemini for diagnosing cases of neuro-ophthalmology. *Cureus*. 2024, 16:e58232. [10.7759/cureus.58232](https://doi.org/10.7759/cureus.58232)
7. Flesch-Kincaid Readability Test. (2007). <https://www.scirp.org/reference/referencespapers?referenceid=3343013>.
8. Fitria TN: QuillBot as an online tool: students' alternative in paraphrasing and rewriting of English writing . *Englisia J*. 2021, 9:1.
9. Charnock D, Shepperd S, Needham G, Gann R: DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. *J Epidemiol Community Health*. 1999, 53:105-11. [10.1136/jech.53.2.105](https://doi.org/10.1136/jech.53.2.105)
10. Kirchner GJ, Kim RY, Weddle JB, Bible JE: Can artificial intelligence improve the readability of patient education materials?. *Clin Orthop Relat Res*. 2023, 481:2260-7. [10.1097/CORR.0000000000002668](https://doi.org/10.1097/CORR.0000000000002668)
11. Patel EA, Fleischer L, Filip P, et al.: The use of artificial intelligence to improve readability of otolaryngology patient education materials. *Otolaryngol Head Neck Surg*. 2024, 171:603-8. [10.1002/ohn.816](https://doi.org/10.1002/ohn.816)
12. Rouhi AD, Ghanem YK, Yolchieva L, et al.: Can artificial intelligence improve the readability of patient education materials on aortic stenosis? A pilot study. *Cardiol Ther*. 2024, 13:137-47. [10.1007/s40119-023-00347-0](https://doi.org/10.1007/s40119-023-00347-0)
13. Roberts H, Zhang D, Dyer GS: The readability of AAOS patient education materials: evaluating the progress since 2008. *J Bone Joint Surg Am*. 2016, 98:e70. [10.2106/JBJS.15.00658](https://doi.org/10.2106/JBJS.15.00658)
14. Thorat V, Rao P, Joshi N, Talreja P, Shetty AR: Role of artificial intelligence (AI) in patient education and communication in dentistry. *Cureus*. 2024, 16:e59799. [10.7759/cureus.59799](https://doi.org/10.7759/cureus.59799)
15. Behers BJ, Vargas IA, Behers BM, Rosario MA, Wojtas CN, Deevers AC, Hamad KM: Assessing the readability of patient education materials on cardiac catheterization from artificial intelligence chatbots: an observational cross-sectional study. *Cureus*. 2024, 16:e63865. [10.7759/cureus.63865](https://doi.org/10.7759/cureus.63865)
16. Mishra K, Dabas A: Publication ethics. *Indian Pediatr*. 2021, 58:781-5. [10.1007/s13312-021-2291-6](https://doi.org/10.1007/s13312-021-2291-6)
17. 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, 14:482-6. [10.4103/idoj.idoj_72_23](https://doi.org/10.4103/idoj.idoj_72_23)
18. Yüce A, Erkurt N, Yerli M, Misir A: The potential of ChatGPT for high-quality information in patient education for sports surgery. *Cureus*. 2024, 16:e58874. [10.7759/cureus.58874](https://doi.org/10.7759/cureus.58874)