Epilepsy & Behavior 163 (2025) 110193

Contents lists available at ScienceDirect

Epilepsy & Behavior

journal homepage: www.elsevier.com/locate/yebeh

Can people with epilepsy trust AI chatbots for information on physical exercise?

Rizia Rocha-Silva [a,b], Bráulio Evangelista de Lima [b], Thalles Guilarducci Costa [b,c], Naiane Silva Morais [b], Geovana José [d], Douglas Farias Cordeiro [d], Alexandre Aparecido de Almeida [e], Glauber Menezes Lopim [f], Ricardo Borges Viana [g], Bolivar Saldanha Sousa [h], Diego Basile Colugnati [i], Rodrigo Luiz Vancini [j], Marília Santos Andrade [k], Katja Weiss [l], Beat Knechtle [l,m,*], Ricardo Mario Arida [k], Claudio Andre Barbosa de Lira [b,n]

[a] Center for Teaching and Research Applied to Education, Federal University of Goiás, Goiânia, Brazil
[b] Faculty of Medicine, Postgraduate Program in Health Sciences, Federal University of Goiás, Goiânia, Brazil
[c] State University of Goias, University Unit of Itumbiara, Itumbiara, Goiás, Brazil.
[d] Faculty of Information and Communication, Federal University of Goiás, Goiânia, Brazil
[e] Federal Institute of Education, Science and Technology of Tocantins, Tocantins, Brazil
[f] Faculty of Medicine, Faculty Faceres, São José do Rio Preto, São Paulo, Brazil
[g] Institute of Physical Education and Sports, Federal University of Ceará, Fortaleza, Brazil
[h] Santa Casa de São Paulo School of Medical Sciences, São Paulo, Brazil
[i] Department of Physiological Sciences, Institute of Biological Sciences, Federal University of Goiás, Goiânia, Brazil
[j] Center for Physical Education and Sports, Federal University of Espírito Santo, Vitória, Brazil
[k] Department of Physiology, Federal University of São Paulo, São Paulo, Brazil
[l] Institute of Primary Care, University of Zurich, Zurich, Switzerland
[m] Medbase St. Gallen Am Vadianplatz, Switzerland
[n] Faculty of Physical Education and Dance, Federal University of Goiás, Goiânia, Brazil

ARTICLE INFO

Keywords:
Epilepsy
Health Information
Physical Exercise
Artificial Intelligence
AI Chatbots

ABSTRACT

Purpose: This study aims to evaluate the similarity, readability, and alignment with current scientific knowledge of responses from AI-based chatbots to common questions about epilepsy and physical exercise.
Methods: Four AI chatbots (ChatGPT-3.5,ChatGPT 4, Google Gemini, and Microsoft Copilot) were evaluated. Fourteen questions on epilepsy and physical exercise were designed to compare the platforms. Lexical similarity, response patterns, and thematic content were analyzed. Readability was measured using the Flesch Reading Ease and Flesch–Kincaid Grade Level scores. Seven experts rated the quality of responses on a Likert scale from "very poor" to "very good."
Results: The responses showed lexical similarity, with approaches to physical exercise ranging from conservative to holistic. Microsoft Copilot scored the highest on the Flesch Reading Ease scale (48.42 ± 13.71), while ChatGPT-3.5 scored the lowest (23.84 ± 8.19). All responses were generally rated as difficult to read. Quality ratings ranged from "Good" to "Acceptable," with ChatGPT 4 being the preferred platform, chosen by 48.98 % of reviewers.

Abbreviations: PWE, People with Epilepsy; AI, Artificial Intelligence; NLP, Natural Language Processing; LLM, Large Language Models; GPT, Generative Pre-trained Transformer; SEO, Search Engine Optimization; JASP, Jeffreys's Amazing Statistics Program; ANOVA, Analysis of Variance; SD, Standard Deviation.

* Corresponding author at: Institute of Primary Care, University of Zurich, Zurich, Switzerland.
E-mail addresses: rizia.rocha@ufg.br (R. Rocha-Silva), brauliofisio@discente.ufg.br (B.E. de Lima), thalles.costa@ueg.br (T.G. Costa), naiane_silva_morais@discente.ufg.br (N.S. Morais), geo_jose@discente.ufg.br (G. José), cordeiro@ufg.br (D.F. Cordeiro), alexandre.almeida@ifto.edu.br (A.A. de Almeida), glaubermenezes@uniterp.com.br (G.M. Lopim), vianaricardoborges@ufc.br (R.B. Viana), bsaldanhas@yahoo.com.br (B.S. Sousa), dcolugnati@ufg.br (D.B. Colugnati), rodrigo.vancini@ufes.br (R.L. Vancini), marilia.andrade@unifesp.br (M.S. Andrade), katja@weiss.co.com (K. Weiss), beat.knechtle@hispeed.ch (B. Knechtle), arida@unifesp.br (R.M. Arida), claudioandre@ufg.br (C.A.B. de Lira).

https://doi.org/10.1016/j.yebeh.2024.110193
Received 30 May 2024; Received in revised form 18 November 2024; Accepted 27 November 2024
Available online 4 December 2024

Conclusion: The findings highlight the potential of AI chatbots as useful sources of information on epilepsy and physical exercise. However, simplifying language and tailoring content to user's needs is essential to enhance their effectiveness.

1. Introduction

Research has consistently shown the undeniable benefits of physical exercise for people with epilepsy (PWE) [1]. Exercise has been found to improve cardiorespiratory and muscular fitness [2], enhance quality of life [3], boost mood [4], and positively impact body composition [5] in PWE. Furthermore, both human and animal studies suggest that physical exercise may help reduce seizure frequency [6,7], potentially serving as a protective measure for epilepsy [8]. Despite this growing knowledge, a significant number of PWE remain inactive, and surprisingly, many healthcare professionals still hold misconceptions about the safety and benefits of exercise for individuals with epilepsy [9–11].

Over the past 30 years, access to information has expanded rapidly with the rise of new technologies like the Internet and social media [12]. For many, the Internet has become a key source of health information [13,14]. Most recently, advances in artificial intelligence (AI), particularly in deep learning, have allowed machines to learn and improve without direct programming [15]. This progress has also enhanced natural language processing (NLP), enabling computers to better understand and interact with human language [16]. The development of generative AI has further led to the emergence of large language models (LLM) [17], which power modern chatbots and have transformed how people engage with digital information [18,19].

These chatbots provide a more dynamic and interactive experience compared to the static, one-way information flow of traditional websites. They are trained on vast amounts of publicly available text data, including books, articles, and webpages [20]. This allows them to process large datasets and engage in advanced, human-like conversations using natural language [20,21]. With the rise of platforms like Generative Pre-trained Transformer (ChatGPT) 3.5 and the release of ChatGPT 4 in 2023, these technologies have gained widespread recognition. Additionally, other accessible platforms like Microsoft Copilot (formerly Bing Chat) and Google Gemini (formerly Bard) have emerged, further advancing NLP technology [16].

While the Internet has undoubtedly democratized information access, it has also introduced challenges [22]. Research has shown that this increased access can sometimes contribute to the spread of misinformation and misconceptions across various subjects [23,24]. This issue is particularly concerning in the health domain, where the quality of information can have direct consequences on people's lives and well-being [25].

Research on chatbots in healthcare has expanded, offering new opportunities for sharing medical information and supporting patients [26–31]. Epilepsy is one of the most common neurological conditions, affecting over 50 million people worldwide. Physical exercise, a key non-pharmacological approach for preventing and managing non-communicable diseases, including epilepsy [32], underscores the importance of chatbot-provided information being both reliable and accessible to the general public. Additionally, as these recommendations often target clinical audiences, ensuring the safety of the information is crucial. Ensuring that health behavior recommendations are accurate and free from misleading information is particularly important for clinical populations where safety is a primary concern. This study, therefore, aims to evaluate the similarity, readability, and quality of information provided by four AI-integrated chatbots in response to common questions about epilepsy and physical exercise. The quality of the information was assessed based on current physical exercise guidelines.

2. Materials and methods

2.1. Question creation and AI chatbots

The most common questions related to "physical activity and epilepsy," "epilepsy and physical exercise," and "epilepsy and exercise" were identified using Google. To determine the primary questions, we used the keyword research tool keywordtool.io [33], which helps users find relevant keywords for search engine optimization (SEO) by generating suggestions based on real Google searches via Google Autocomplete. The data was collected on December 4, 2023. Additional relevant questions were sourced from epilepsy-related websites [34–38]. In total, 14 questions were formulated: five covered general aspects of epilepsy, and nine focused on the relationship between epilepsy and physical exercise (Table 1).

Four well-known and easily accessible AI chatbots were selected: ChatGPT-3.5 (database updated in January 2022), ChatGPT 4 (database updated in April 2023 and connected to Bing), Gemini (continuously updated and connected to the Internet), and Microsoft Copilot (continuously updated and connected to the Internet). Responses were extracted on December 8, 2023. A new chat was initiated for each question to prevent earlier questions from affecting the answers.

2.2. Qualitative analysis

A similarity analysis was employed to examine word co-occurrence in the responses [39]. This method is a text mining and data visualization technique that maps relationships between terms in a text dataset, showing words, topics, or concepts that frequently appear together in clusters, indicating potential contextual connections. Distinct clusters were created to illustrate the correlations and frequency of association among the terms [40]. We selected 45 terms per chatbot to provide a clear visualization of these relationships, focusing on the most frequently co-occurring words that had meaningful context in the responses to minimize visual clutter and ensure accurate interpretation [41].

2.3. Evaluation of the questions

To thoroughly assess the readability of the responses, the Flesch Reading Ease scale and the Flesch–Kincaid Grade Level were utilized [42] using the Flesch reading ease test score calculator [43]. The Flesch Reading Ease score evaluates how easy a text is to read, ranging from 0 to 100 points. This score reflects the level of difficulty an average adult might experience when understanding the text, with higher scores indicating greater accessibility and clarity. Scores are categorized as very easy to read (90–100 points), easy to read (80–90 points), fairly easy to read (70–80 points), plain English (60–70 points), fairly difficult to read (50–60 points), difficult to read (30–50 points), very difficult to read (10–30 points), and extremely difficult to read (below 10 points) [43].

The Flesch–Kincaid Grade Level score assesses readability based on educational level, indicating the number of years of schooling needed to comprehend the text. This method analyzes factors like sentence length and word count to estimate the required educational level. A lower score indicates that the text is easier to read, with the lowest scores representing the most readable content [42]. Scores are categorized as follows: 5th grade (90–100 points), 6th grade (80–90 points), 7th grade (70–80 points), 8th and 9th grade (60–70 points), 10th–12th grade (50–60 points), College (30–50 points), College graduate (10–30 points), and Professional (below 10 points). Additional readability factors considered included the word count of each response and the number of sentences used to express the ideas.

2.4. Quality evaluation of responses

The quality of the responses was evaluated based on current scientific evidence through a blind review process involving seven expert reviewers with academic expertise in exercise physiology, neuroscience, neurology, and epilepsy (MSA, RLV, AAA, GML, BSS, DBC, and RMA) (for the details see Supplementary Table 1). These reviewers came from diverse educational backgrounds, including exercise and sports science, medicine, physiotherapy, and biomedical science, and all held Ph.D. degrees. They were blinded to any details about the responses or the identities of the chatbots that provided them. The reviewers received the responses via Google Forms (Alphabet, Mountain View CA, USA) and rated the quality on a 5-point Likert scale, where 1 is very poor, 2 is poor, 3 is acceptable, 4 is good, and 5 is very good [28]. Each evaluation considered the accuracy of the information, clarity, and detail. For a response to be rated as "good," it needed to meet all criteria: being accurate, evidence-based, and well-formulated, directly addressing the question, and using accessible language with clearly explained technical terms. More information about the evaluation criteria can be found in Supplementary Table 2. The reviewers also identified which ranking represented the best answer for each question. With 14 questions across 4 chatbots, each reviewer evaluated 56 responses (Supplementary raw data in Microsoft Excel format).

2.5. Statistical analysis

A similarity analysis of words was conducted on the responses generated by AI-integrated chatbots. The textual *corpus* was evaluated by extracting the main co-occurring terms, standardized at 45 terms for each chatbot, revealing communities of similarity. Iramuteq software (Version 0.7, alpha 2, Laboratoire LERASS, France), based on the R language, was used for this analysis [40,44].

Data were analyzed using Jeffreys's Amazing Statistics Program (JASP, 0.18.1.0, Amsterdam University, Netherlands). Normality was assessed with the Shapiro–Wilk test, and Levene's test evaluated the equality of variances. Analysis of variance (ANOVA) was performed to compare differences among the AIs Gemini, ChatGPT-3.5, ChatGPT 4, and Microsoft Copilot for the following variables: word count, number of sentences, Flesch Reading Ease score, and Flesch–Kincaid Grade Level. Partial eta squared (η_p^2) was calculated as the effect size [45], following the benchmarks proposed by Espírito-Santo and Daniel [46]: "small" ($0.001 \leq \eta_p^2 < 0.06$), "medium" ($0.06 \leq \eta_p^2 < 0.14$), and "large" ($\eta_p^2 \geq 0.14$). Due to the heterogeneity of the data, Welch correction was applied, and post-hoc analysis was performed when necessary [47]. Data were expressed as mean and standard deviation (SD), along with minimum and maximum values.

The bootstrapping method (1,000 resamples) was employed to enhance the reliability of the results and correct for normality deviations when needed (number of sentences). A 95 % confidence interval (CI) for the bootstrapping was reported [48]. Welch's correction was applied for variance heterogeneity when required (Flesch Reading Ease score and Flesch–Kincaid Grade Level). A significance level of 5 % was used for statistical analyses.

3. Results

3.1. Qualitative analyses

The similarity of words in the responses from each chatbot was examined (Fig. 1). All chatbots exhibited three central themes: "epilepsy," "seizure," and "exercise," with similarities in how their clusters were constructed and in word co-occurrence patterns. The strength of the relationship between the central terms is indicated by the width of the edges connecting the clusters, with wider edges indicating a stronger thematic relationship.

Overall, Microsoft Copilot emphasizes safety and control in relation to "seizure" (Fig. 1A) and, along with ChatGPT 4, features a cluster linking "provider" and "healthcare," highlighting the interaction of individuals with epilepsy in a medical context (Fig. 1B). ChatGPT-3.5 is unique in placing "seizure" within a broader context of "epilepsy" and presents fewer clusters, suggesting a more limited range of themes discussed (Fig. 1D).

All chatbots addressed physical exercise from different viewpoints. Microsoft Copilot emphasizes exercise as a treatment method, taking a more conservative stance by highlighting terms like "help" and "regimen." Its connections to other clusters are represented by thinner edges, which may indicate a weaker or less direct relationship (Fig. 1A). In contrast, the thicker edges in ChatGPT 4 indicate a strong connection between "exercise" and "seizure," as well as between "epilepsy" and "individual," underscoring the importance of physical exercise and seizure management in the lives of PWE and the need for personalized treatment (Fig. 1B).

Gemini associated the physical exercise cluster with terms that suggest a careful approach, including "regular," "doctor," "health," "body," and "mood." This indicates that physical exercise is a vital component of health management for PWE, highlighting the importance of regularity and tailoring intensity to the individual's health status (Fig. 1C). Meanwhile, the relationships between concepts in ChatGPT-3.5 show a thought consideration of how exercise and physical activity relate to the condition of individuals with epilepsy, emphasizing safety, types of exercises, and the need for supervision by a healthcare professional (Fig. 1D).

Table 1
Questions related to epilepsy and physical exercise.

Questions
1. Are epilepsy and seizures the same?
2. How does epilepsy work?
3. What are the causes of epilepsy?
4. Is it possible to live a normal life with epilepsy?
5. Can epilepsy be cured?
6. Is physical exercise beneficial for epilepsy?
7. Can individuals with epilepsy engage in physical exercise?
8. How should someone with epilepsy begin exercising?
9. How does epilepsy impact physical activity participation?
10. Can physical exercise improve epilepsy?
11. Can physical exercise trigger seizures in people with epilepsy?
12. Is it safe to exercise after a seizure?
13. What are the best types of exercise for individuals with epilepsy?
14. Can people with epilepsy perform high-intensity exercises?

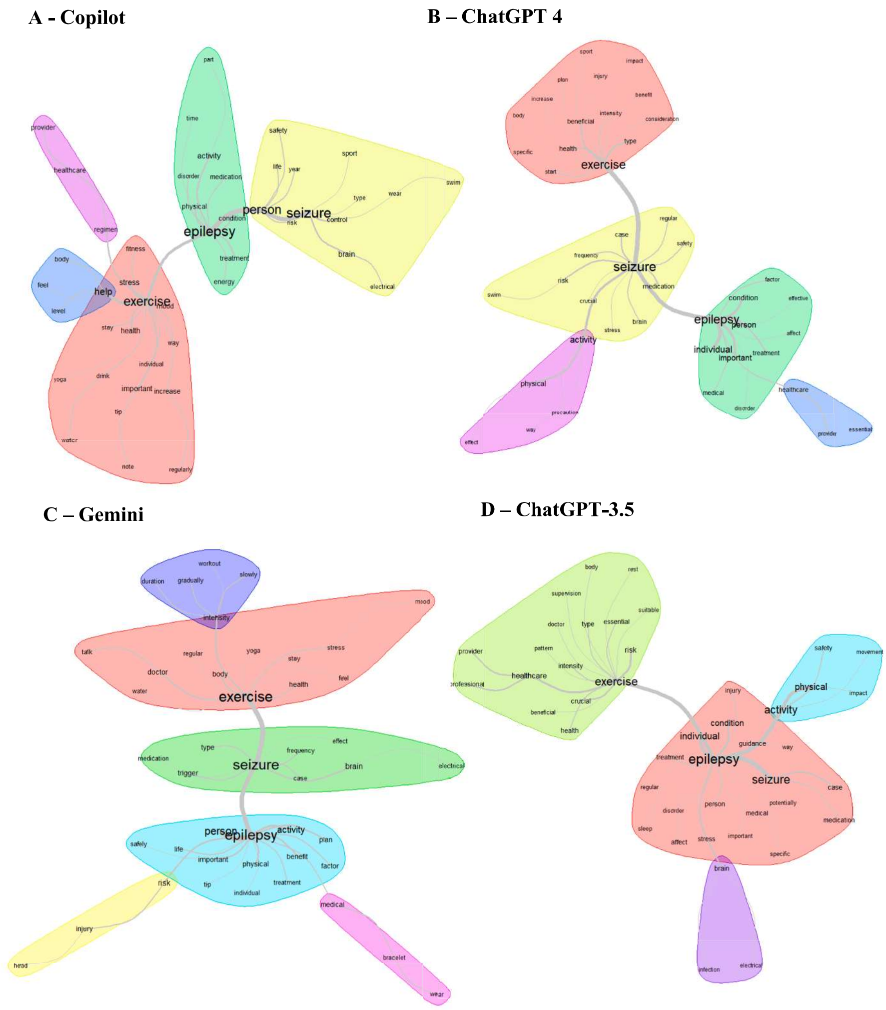

Fig. 1. Similarity graph based on the co-occurrence of responses generated by AI-integrated chatbots. The colors represent the default settings of Iramuteq software and are used to visually differentiate word clusters.

3.2. Evaluation of the questions

3.2.1. Word count per response

The average word count for each response was 192.07 (SD 52.23; Min-Max 82.00–292.00) for Microsoft Copilot, 311.36 (SD 74.94; Min-Max 148.00–427.00) for ChatGPT 4, 305.21 (SD 99.88; Min-Max 179.00–478.00) for Gemini, and 176.71 (SD 76.43; Min-Max 78.00–381.00) for ChatGPT-3.5. The normality test indicated that the word count variable follows a normal distribution (Shapiro–Wilk $p > 0.05$). Levene's test confirmed equal variances across the groups ($F [3,52] = 2.09$; $p = 0.113$). ANOVA indicated a significant difference among the groups ($F [3,52] = 11.965$; $p < 0.001$; $\eta_p^2 = 0.408$). Microsoft Copilot had a lower word count compared to ChatGPT 4 ($p < 0.001$) and Gemini ($p = 0.002$). The descriptive results from the post-hoc analysis with Bonferroni correction are shown in Table 2.

3.2.2. Number of sentences

The average number of sentences per response was 12.07 (SD 7.89; Min-Max 4.00–33.00) for ChatGPT-3.5, 14.00 (SD 5.24; Min-Max 6.00–22.00) for Microsoft Copilot, 18.00 (SD 6.20; Min-Max 11.00–33.00) for Gemini, and 23.42 (SD 5.80; Min-Max 11.00–35.00) for ChatGPT 4. The normality test showed that sentence lengths did not follow a normal distribution for Gemini (Shapiro–Wilk = 0.834; $p = 0.014$) and ChatGPT-3.5 (Shapiro–Wilk = 0.838; $p = 0.015$). Levene's

Table 2
Post-hoc results with Bonferroni correction comparing the AI chatbots Microsoft Copilot, ChatGPT-3.5,ChatGPT 4, and Gemini regarding word count in response to epilepsy questions.

Variable comparison		Δ	SE	95 % CI	t value	Cohen's d (95 %IC)	P_{bonf}
Copilot	ChatGPT-4	−119.28	29.37	−197.25 to −41.32	−4.061	−1.535 (−2.651; −0.419)	< 0.001*
	Gemini	−113.14	29.37	−191.10 to −35.17	−3.852	−1.456 (−2.564; −0.348)	0.002*
	ChatGPT-3.5	15.35	29.37	−62.60 to 93.32	0.523	0.198 (−0.841; 1.236)	1.000
ChatGPT-4	Gemini	6.14	29.37	−71.82 to 84.10	0.209	0.079 (−0.958; 1.116)	1.000
	ChatGPT-3.5	134.64	29.37	56.67 to 212.60	4.584	1.732 (0.596; 2.869)	< 0.001*
Gemini	ChatGPT-3.5	128.50	29.37	50.53 to 206.46	4.374	1.653 (0.525; 2.781)	< 0.001*

Δ, mean difference; SE, standard error; 95% CI, confidence interval; P_{bonf}, *p*-value from Bonferroni correction from the post-hoc analysis.
* Statistical significance at 0.05.

test revealed that the variances among the groups were equal (F [3,52] = 0.596; $p = 0.620$). ANOVA found significant differences between the groups (F [3,52] = 8.704; $p < 0.001$; $\eta_p^2 = 0.334$), and the descriptive results from the post-hoc analysis, based on bootstrapping (1,000 successful resamples), are shown in Table 3.

3.2.3. Flesch Reading Ease score

The Flesch Reading Ease scores for the chatbots were 48.42 (SD 13.71; Min-Max 14.12–63.49) for Microsoft Copilot, 35.46 (SD 9.58; Min-Max 21.03–53.72) for ChatGPT 4, 38.94 (SD 11.93; Min-Max 16.15–57.81) for Gemini, and 23.84 (SD 8.19; Min-Max 12.29–41.90) for ChatGPT-3.5. The normality test indicated that the Flesch Reading Ease scores were normally distributed ($p > 0.05$). Levene's test confirmed equal variances among the groups (F [3,52] = 1.25; $p = 0.299$). ANOVA revealed significant differences between the groups (Welch's F [3,52] = 11.800; $p < 0.001$; $\eta_p^2 = 0.405$). Although Microsoft Copilot had the highest average readability scores ($p = 0.019$), significant differences were found among the chatbots, with ChatGPT-3.5 showing the lowest readability ($p < 0.001$). The scores fell within the range of "Very difficult to read" to "Difficult to read." The post-hoc results with Bonferroni correction are detailed in Table 4.

3.2.4. Flesch–Kincaid Grade Level score

The Flesch–Kincaid Grade Level scores for the chatbots were 14.03 points (SD 1.91; Min-Max 10.28–17.13) for ChatGPT-3.5, 12.11 points (SD 1.80; Min-Max 8.81–14.57) for Gemini, 11.63 points (SD 1.47; Min-Max 8.96–13.79) for ChatGPT 4, and 10.03 points (SD 2.49; Min-Max 7.03–15.68) for Microsoft Copilot. The normality test indicated that the grade level scores were normally distributed ($p > 0.05$). Levene's test confirmed equal variances among the groups (F [3,52] = 1.35; $p = 0.267$). ANOVA revealed significant differences between the groups (Welch's F [3,52] = 9.916; $p < 0.001$; $\eta_p^2 = 0.364$). The post-hoc results with Bonferroni correction are detailed in Table 5.

3.3. Quality evaluation

Table 6 presents the quality assessments of the chatbots conducted by experts. ChatGPT 4 was the preferred option among reviewers, receiving 48.98 % of the votes. The quality ratings varied across different research sources. Microsoft Copilot and ChatGPT 4 were primarily rated as "Good," each receiving 71.43 % (10 out of 14) of assessments in that category. Gemini was rated as "Good" for 50.00 % (7 out of 14) of its responses, while ChatGPT-3.5 had 64.29 % (9 out of 14) rated as "Acceptable."

4. Discussion

This study evaluated the similarity, readability, and quality of information provided by four AI-integrated chatbots in response to common questions about epilepsy and physical exercise, specifically focusing on whether the information aligns with current exercise guidelines. The results reveal that the chatbots exhibited similar response patterns based on word co-occurrence, offering answers that ranged from "good" to "acceptable." ChatGPT 4 was found to provide the best responses according to expert evaluations. However, texts generated by these tools demonstrated low readability.

This study is relevant as epilepsy is a chronic, non-communicable disease affecting approximately 50 million people globally. Moreover, 80 % of PWE reside in low- and middle-income countries, and 75 % of those in low-income countries [49] do not receive adequate treatment due to limited access to healthcare professionals and specialized services. In this context, these individuals may seek information about their condition online and through AI tools. Research indicates that people with chronic illnesses are more likely to use internet search engines, like Google, to find information about medications, symptoms, and related topics [50]. This is especially true for PWE, who require daily self-management and often turn to the Internet to better understand their condition [50].

The rise of AI tools has led to broader access to information, including health-related information. Chatbots, featuring a user-friendly, chat-based interface, are designed to comprehend and respond to questions in natural language. They have greatly improved communication by overcoming geographical and linguistic barriers [51,52].

In this study, the evaluated chatbots exhibited similarities in their response structures, as indicated by the patterns of word co-occurrence clusters, which may be influenced by their data sources, usage policies, and processing models. For example, ChatGPT-3.5 tends to provide more general responses and has fewer clusters, which may restrict the depth of information provided. In contrast, ChatGPT 4 and Google's Gemini focus on personalizing advice related to physical exercise. Although Microsoft Copilot is based on ChatGPT 4, it offers more conservative information regarding physical exercise, lacking details about exercise intensity or types of activities.

Reading comprehension is a vital skill for sharing health-related information [53]. Medical texts that exceed patients' reading abilities can lead to adverse health outcomes [53]. Thus, we evaluated the readability of responses provided by chatbots. The scores ranged from "Very Difficult to Read" to "Difficult to Read," indicating a requirement for higher education levels. Our analysis found that many responses contained lengthy texts and long sentences, suggesting that a more advanced reading ability [42] and education are needed for full understanding. As a result, the findings indicate low readability levels. Given that 80 % of PWE live in low- and middle-income countries [54], where educational attainment is generally lower than in high-income countries, these low readability levels may hinder access to information from AI tools. Microsoft Copilot was one of the chatbots with higher readability, utilizing short sentences and fewer words in its responses.

While some websites provide information on specific topics related to epilepsy, many PWE find that the excessive content can be overwhelming and that there is often no adequate search tool to filter the necessary information [55]. In contrast, AI tools like chatbots offer the advantage of personalizing content, tailoring responses to meet the user's needs and context. For example, if a user finds a response difficult

Table 3
Post-hoc results from bootstrapping (1,000 resamples) comparing AI chatbots Microsoft Copilot, ChatGPT-3.5,ChatGPT 4, and Gemini regarding the number of sentences in response to epilepsy questions.

Variable comparison		Δ	SE	95 % CI	t value	Cohen's d (95 %IC)	P_{bonf}
Copilot	ChatGPT-4	−9.49	2.07	−13.349 to −5.217	−3.919	−1.481 (−2.592; 0.371)	0.002*
	Gemini	−3.91	2.22	−9.250 to −0.179	−1.663	−0.628 (−1.679; 0.422)	0.614
	ChatGPT-3.5	2.08	2.48	−3.780 to 5.924	0.802	0.303 (−0.737; 1.343)	1.000
ChatGPT-4	Gemini	5.46	2.25	0.675 to 9.615	2.257	0.853 (−0.209; 1.915)	0.170
	ChatGPT-3.5	11.56	2.50	5.767 to 15.338	4.721	1.784 (0.642; 2.927)	< 0.001*
Gemini	ChatGPT-3.5	5.91	2.70	0.512 to 11.025	2.465	0.932 (−0.135; 1.998)	0.102

Δ, the average difference estimate is derived from the median of the bootstrap distribution; SE, standard error; 95% CI † bca, confidence interval for the marginal means based on 1,000 bootstrap resamples; Pbonf, p-value from the Bonferroni correction from the post-hoc analysis via bootstrapping.
* Statistical significance at the 0.05 level.

Table 4
Post-hoc results with Bonferroni correction comparing AI chatbots Microsoft Copilot, ChatGPT-3.5,ChatGPT 4, and Gemini regarding Flesch Reading Ease scores in response to epilepsy questions.

Variable comparison		Δ	SE	95 % CI	t value	Cohen's d (95 %IC)	P_{bonf}
Copilot	ChatGPT-4	12.95	4.18	−0.98 to 26.90	3.099	1.171 (0.088; 2.255)	0.019*
	Gemini	9.48	4.18	−4.45 to 23.42	2.269	0.857 (−0.205; 1.920)	0.165
	ChatGPT-3.5	24.58	4.18	10.65 to 38.52	5.880	2.223 (1.026; 3.419)	< 0.001*
ChatGPT-4	Gemini	−3.47	4.18	−17.41 to 10.46	−0.831	−0.314 (−1.354; 0.726)	1.000
	ChatGPT-3.5	11.62	4.18	−2.31 to 25.57	2.781	1.051 (−0.023; 2.126)	0.045*
Gemini	ChatGPT-3.5	15.10	4.18	1.17 to 29.04	3.612	1.365 (0.265; 2.465)	0.004*

Δ: mean difference; SE: Standard error; 95%CI: Confidence interval; Pbonf: p-value from Bonferroni correction of the post-hoc.
* Statistical significance at 0.05 level.

Table 5
Post-hoc results with Bonferroni correction comparing AI chatbots Microsoft Copilot, ChatGPT-3.5,ChatGPT 4, and Gemini regarding grade level scores in response to epilepsy questions.

Variable comparison		Δ	SE	95 % CI	t value	Cohen's d (95 %IC)	P_{bonf}
Copilot	ChatGPT-4	−1.59	0.74	−-3.563 to 0.364	−2.161	−0.817 (−1.877; 0.243)	0.212
	Gemini	−2.08	0.74	−4.047 to −0.120	−2.816	−1.064 (−2.140; 0.011)	0.041*
	ChatGPT-3.5	−4.00	0.74	−5.964 to −2.036	−5.406	−2.043 (−3.217; −0.870)	< 0.001*
ChatGPT-4	Gemini	−0.48	0.74	−2.448 to 1.479	−0.655	−0.247 (−1.286; 0.791)	1.000
	ChatGPT-3.5	−2.40	0.74	−4.364 to −0.437	−3.245	−1.226 (−2.314; −0.138)	0.012*
Gemini	ChatGPT-3.5	−1.91	0.74	−3.880 to 0.047	−2.590	−0.979 (−2.049; 0.091)	0.075

Δ, mean difference; SE, standard error; 95% CI, confidence interval; Pbonf, *p*-value from the Bonferroni correction from the post-hoc analysis.
* Statistical significance at the 0.05 level.

Table 6
Assessment of AI-integrated chatbot quality.

Questions	**Sources**				**Reviewer preference**
	Microsoft Copilot (%)	**ChatGPT-4** (%)	**Gemini Google** (%)	**ChatGPT 3.5** (%)	
1. Are epilepsy and seizures the same?	Acceptable (42.85) Good (42.85)	Good (57.14)	Acceptable (42.85)	Very Good (57.14)	ChatGPT 3.5
2. How does epilepsy work?	Acceptable (42.85) Good (42.85)	Good (57.14)	Good (57.14)	Acceptable (57.14)	ChatGPT-4
3. What are the causes of epilepsy?	Acceptable (42.85) Good (42.85)	Good (42.85)	Acceptable (42.85) Good (42.85)	Poor (42.85) Acceptable (42.85)	ChatGPT-4
4. Is it possible to live a normal life with epilepsy?	Good (71.42)	Good (71.42)	Acceptable (42.85)	Poor (57.14)	Gemini
5. Can epilepsy be cured?	Acceptable (42.85)	Very good(57.14)	Acceptable (42.85) Good (42.85)	Acceptable (71.42)	ChatGPT-4
6. Is physical exercise beneficial for epilepsy?	Good (71.42)	Good (42.85)	Acceptable (42.85)	Acceptable (42.85)	Gemini
7. Can individuals with epilepsy engage in physical exercise?	Acceptable (57.14)	Good (57.14)	Good (57.14)	Good (57.14)	ChatGPT-4
8. How should someone with epilepsy begin exercising?	Good (71.42)	Very good (42.85) Good (42.85)	Very good (42.85)	Acceptable (42.85)	Gemini
9. How does epilepsy impact physical activity participation?	Good (71.42)	Very good(42.85)	Good (42.85)	Acceptable (42.85) Good (42.85)	ChatGPT-4
10. Can physical exercise improve epilepsy?	Good (71.42)	Good (71.42)	Very good (57.14)	Good (42.85)	ChatGPT-4
11. Can physical exercise trigger seizures in people with epilepsy?	Acceptable (42.85) Good (42.85)	Good (57.14)	Acceptable (42.85)	Acceptable (42.85)	ChatGPT-4
12. Is it safe to exercise after a seizure?	Poor (57.14)	Good (71.42)	Good (57.14)	Acceptable (42.85)	ChatGPT-4
13. What are the best types of exercise for individuals with epilepsy?	Acceptable (57.14)	Very good (57.14)	Good (57.14)	Acceptable (42.85)	ChatGPT-4
14. Can people with epilepsy perform high-intensity exercises?	Good (57.14)	Very good (57.14)	Very good (42.85)	Good (85.71)	ChatGPT-4

to understand, they can indicate their confusion, prompting the chatbot to modify the response to make it clearer and more comprehensible [25].

In evaluating the content quality of the responses, most ratings from reviewers were categorized as “good” or “acceptable,” with Google Gemini andChatGPT 4 receiving “very good” ratings. These findings align with those of Wu et al. [19], who evaluated the responses from ChatGPT-3.5 to 378 epilepsy-related questions, verified by 3 reviewers. They found that 68.4 % of ChatGPT-3.5’s responses were “correct and comprehensive,” making it a valuable source of reliable information for educating PWE. However, it was noted that ChatGPT-3.5 struggled to provide prognostic information and is not recommended for medical guidance. In a separate study by Kim et al. [56], which compared ChatGPT-3.5 and ChatGPT 4 on 57 common epilepsy questions, it was concluded that 40 of those questions had “sufficient educational value,” with ChatGPT 4 demonstrating its potential as a reliable information source on epilepsy.

In one of the limited studies comparing ChatGPT-3.5, Gemini, and Bing, Salazar et al. [21] assessed their accuracy in distinguishing between medical emergencies and non-emergencies. The chatbots identified 12–15 % more cases as emergencies than the reviewers did. They concluded that while these systems need further improvement for emergency detection, they have the potential to be valuable tools in patient care and support.

Chatbots do not always deliver definitive or specialized information, as their AI training emphasizes general cognitive skills rather than focusing exclusively on medical or health applications [56]. This was evident in the responses in the current study. In response to the sixth question regarding whether physical exercise is beneficial for epilepsy, ChatGPT 4 mentioned the interaction between physical exercise and antiepileptic drugs, suggesting that exercise could modify their effects, a claim not backed by scientific evidence [57]. While there may be minor changes in the serum levels of first-generation medications (such as phenytoin and phenobarbital) before and after exercise, these changes are not statistically significant [32,58]. Additionally, no response regarding the causes of epilepsy aligned with the ILAE’s 2017 position statement document [59], which may lead to confusion for readers trying to classify the causes of epilepsy. Another notable point for evaluators was the response to the question “Can epilepsy be cured?” where ChatGPT 4 placed the word “cure” in quotes, which could create confusion by implying that the “cure” depends on certain factors. For a layperson or someone with limited knowledge, this might lead to a misunderstanding that epilepsy can be cured, overlooking the quotation marks around the word.

Other drawbacks include the absence of real-time data updates in ChatGPT-3.5, a feature that ChatGPT 4 addresses by utilizing the Bing web search engine for external results. Similarly, the Gemini and Microsoft Copilot platforms are integrated with the Internet. The lack of references continues to be a point of contention in generative AI, with ChatGPT-3.5 remaining particularly vulnerable to this issue [28]. In contrast, chatbots like Gemini and ChatGPT 4 can provide references upon request, while Microsoft Copilot automatically links responses to frequently searched web pages, organization sites, and scientific articles.

Reviewers also observed that the responses from the chatbots emphasized the importance of a doctor’s role in starting a physical exercise program for PWE. However, there was no mention of the involvement of a physical education or sports professional experienced in creating physical exercise programs.

4.1. Strengths and study limitations

This study has several limitations. Although predefined parameters were established for each classification, the subjective nature of assessing response quality can lead to variations based on individual interpretation. This indicates a need for more objective or standardized criteria for evaluating content quality such as algorithm for supplying ratings. Nevertheless, reducing complex medical information to a five-point Likert scale has inherent limitations in capturing ’quality’ fully. To mitigate subjectivity, we invited professionals with expertise in the subject to evaluate each response. However, a notable limitation is the lack of evaluation by an epileptologist, which could have provided a more clinical viewpoint on the accuracy and relevance of the chatbot responses concerning epilepsy and physical exercise. Additionally, these professionals did not have prior access to the evaluation questions, responses, or the search mechanism employed.

Second, while we referenced commonly asked questions on Google regarding physical exercise and epilepsy, along with structured questions from specialized websites aimed at individuals with epilepsy, this method may not completely represent the perspectives of actual patients and could introduce selection bias. Furthermore, the questions used are limited and may not encompass the full range of concerns that individuals with epilepsy may have about these topics. Another limitation is that, although we conducted text mining on the chatbot responses, we did not perform a detailed content analysis, which could have provided a deeper understanding of the context and nuances within the responses. Future studies could address this gap, enabling a more comprehensive evaluation of the information. Additionally, further research should explore how these recommendations can effectively promote active behavior and adherence to physical activity.

Among the strengths, this study is the first to evaluate responses from various chatbots (ChatGPT-3.5,ChatGPT 4, Google Gemini, and Microsoft Copilot) regarding common questions about epilepsy and physical exercise, thereby broadening the scope of the responses. Another key point is the application of text mining for semantic analysis, which assesses the nature of the responses from each chatbot on the topic. Additionally, the potential of AI as an important educational resource for managing epilepsy is significant. It is important to note that chatbots are meant to be complementary tools and do not replace the guidance of medical professionals and other healthcare providers. They represent a promising solution for enhancing health literacy, especially in low- and middle-income countries where access to healthcare is often limited and specialized professionals are few [60].

5. Conclusion

The comparison of chatbots (ChatGPT-3.5,ChatGPT 4, Google Gemini, and Microsoft Copilot) regarding epilepsy and physical exercise underscores their potential in delivering health information. While there is a consistent knowledge base among the platforms, variations in exercise recommendations necessitate caution in their use. The readability of responses indicates a need for simplification to enhance accessibility, as all chatbots produce texts that are generally difficult for the public to understand. The responses, mostly rated as “Good” to “Acceptable,” particularly from ChatGPT 4, demonstrate the potential of AI in health education, suggesting that further technological advancements could enhance the effectiveness and intuitiveness of these tools.

Statement on the use of artificial intelligence

Chatbots were employed to generate answers to the study questions and assist with translation and spell-checking.

Ethical approval and consent to participate

Since all responses from the chatbots were publicly available and the study did not involve human subjects, obtaining informed consent or approval from an institutional ethics committee was not necessary.

Author contributions

RRS and BEL contributed to the conception and design of the study, data acquisition, analysis, drafting the article, and critical revisions.

TGC, GJ, DFC, and NSM were involved in data acquisition, analysis, data interpretation, and critical revisions of the article. AAA, GML, RBV, BSS, DBC, RLV, MSA, and RMA focused on data interpretation and critical revisions. KW and BK were responsible for critical revisions of the article. CABL played a key role in the study's conception and design, analysis, data interpretation, drafting, and critical revisions of the. All authors contributed to and approved the final manuscript.

CRediT authorship contribution statement

Rizia Rocha-Silva: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Bráulio Evangelista de Lima:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Thalles Guilarducci Costa:** Writing – review & editing, Formal analysis, Data curation. **Naiane Silva Morais:** Writing – review & editing, Formal analysis, Data curation. **Geovana José:** Writing – review & editing, Formal analysis, Data curation. **Douglas Farias Cordeiro:** Writing – review & editing, Formal analysis, Data curation. **Alexandre Aparecido de Almeida:** Writing – review & editing, Data curation. **Glauber Menezes Lopim:** Writing – review & editing, Data curation. **Ricardo Borges Viana:** Writing – review & editing, Data curation. **Bolivar Saldanha Sousa:** Writing – review & editing, Data curation. **Diego Basile Colugnati:** Writing – review & editing, Data curation. **Rodrigo Luiz Vancini:** Writing – review & editing, Data curation. **Marília Santos Andrade:** Writing – review & editing, Data curation. **Katja Weiss:** Writing – review & editing. **Beat Knechtle:** Writing – review & editing. **Ricardo Mario Arida:** Writing – review & editing, Data curation. **Claudio Andre Barbosa de Lira:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Data curation, Conceptualization.

Funding

This research was partially funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) under Finance Code 001. RRS, BEL, TGC, and NSM hold fellowships at CAPES. CABL, MSA, and RMA have productivity fellowships from the Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq #302689/2022-2), the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP #2022/10696-1), and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES-PRINT #88881.310490/2018-01). RLV has a productivity fellowship from the Fundação de Amparo à Pesquisa do Estado do Espírito Santo – Brasil (FAPES).

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rizia Rocha-Silva reports financial support was provided by Coordination of Higher Education Personnel Improvement. Claudio Andre Barbosa de Lira reports financial support was provided by National Council for Scientific and Technological Development. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yebeh.2024.110193.

References

[1] Capovilla G, Kaufman KR, Perucca E, Moshé SL, Arida RM. Epilepsy, seizures, physical exercise, and sports: a report from the ILAE Task Force on Sports and Epilepsy. Epilepsia 2016;57:6–12. https://doi.org/10.1111/epi.13261.

[2] Vancini RL, de Lira CAB, Scorza FA, de Albuquerque M, Sousa BS, de Lima C, et al. Cardiorespiratory and electroencephalographic responses to exhaustive acute physical exercise in people with temporal lobe epilepsy. Epilepsy Behav 2010;19: 504–8. https://doi.org/10.1016/J.YEBEH.2010.09.007.

[3] Arida RM, Cavalheiro EA, da Silva AC, Scorza FA. Physical activity and epilepsy. Sport Med 2008;38:607–15. https://doi.org/10.2165/00007256-200838070-00006.

[4] Vancini RL, de Lira CAB, dos Santos AM, de Lima C, Arida RM. Low levels of maximal aerobic power impair the profile of mood state in individuals with temporal lobe epilepsy. Arq Neuropsiquiatr 2015;73:7–11. https://doi.org/10.1590/0004-282X20140188.

[5] de Lima C, Vancini RL, Arida RM, Guilhoto LMFF, de Mello MT, Barreto AT, et al. Physiological and electroencephalographic responses to acute exhaustive physical exercise in people with juvenile myoclonic epilepsy. Epilepsy Behav 2011;22: 718–22. https://doi.org/10.1016/J.YEBEH.2011.08.033.

[6] Peixinho-Pena LF, Fernandes J, de Almeida AA, Gomes FGN, Cassilhas R, Venancio DP, et al. A strength exercise program in rats with epilepsy is protective against seizures. Epilepsy Behav 2012;25:323–8. https://doi.org/10.1016/j.yebeh.2012.08.011.

[7] Arida RM, Scorza FA, Terra VC, Cysneiros RM, Cavalheiro EA. Physical exercise in rats with epilepsy is protective against seizures: evidence of animal studies. Arq Neuropsiquiatr 2009;67:1013–6. https://doi.org/10.1590/S0004-282X2009000600010.

[8] Van Den Bongard F, Hamer HM, Sassen R, Reinsberger C. Sport and Physical activity in epilepsy: a systematic review. Dtsch Arztebl Int 2020;117:1. https://doi.org/10.3238/ARZTEBL.2020.0001.

[9] Carter JM, McGrew C. Seizure disorders and exercise/sports participation. Curr Sports Med Rep 2021;20:26–30. https://doi.org/10.1249/JSR.0000000000000799.

[10] Carrizosa-Moog J, Ladino LD, Benjumea-Cuartas V, Orozco-Hernández JP, Castrillón-Velilla DM, Rizvi S, et al. Epilepsy, physical activity and sports: a narrative review. Can J Neurol Sci 2018;45:624–32. https://doi.org/10.1017/CJN.2018.340.

[11] Arida RM, Sales EP do N, Teixeira-Machado L, Prado GF do, Gutierre RC, Carrizosa J. Neurologists' knowledge of and attitudes toward physical exercise for people with epilepsy in Latin America. Epilepsy Behav 2022;131:108705. doi: 10.1016/J.YEBEH.2022.108705.

[12] González-Pérez Y. Disinformation in the information age. Eur J Hosp Pharm 2020; 27:319. https://doi.org/10.1136/EJHPHARM-2020-002213.

[13] Chu JT, Wang MP, Shen C, Viswanath K, Lam TH, Chan SSC. How, when and why people seek health information online: qualitative study in Hong Kong. Interact J Med Res 2017;6:e24.

[14] Viana RB, Neves-Silva AVS, Santos DAT, Vancini RL, Andrade MS, Teixeira CVLS, et al. The usage of, and confidence in, social media as study sources among undergraduate students: a cross-sectional survey comparing it with traditional study sources. Educ Inf Technol 2021;26:2233–52. https://doi.org/10.1007/S10639-020-10357-3/TABLES/5.

[15] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electron Mark 2021;31:685–95. https://doi.org/10.1007/S12525-021-00475-2/TABLES/2.

[16] Fanni SC, Febi M, Aghakhanyan G, Neri E. Nat Lang Process 2023:87–99. https://doi.org/10.1007/978-3-031-25928-9_5.

[17] Banerjee S, Dunn P, Conard S, Ng R. Large language modeling and classical AI methods for the future of healthcare. J Med Surgery, Public Heal 2023;1:100026. https://doi.org/10.1016/J.GLMEDI.2023.100026.

[18] Euchner J, Generative AI. Res Technol Manag 2023;66:71–4. https://doi.org/10.1080/08956308.2023.2188861/ASSET//CMS/ASSET/8BB230B2-0BFB-4E02-BF1B-2E9E0C7060CD/08956308.2023.2188861.FP.PNG.

[19] Banh L, Strobel G. Generative artificial intelligence. Electron Mark 2023;33:1–17. https://doi.org/10.1007/S12525-023-00680-1.

[20] Meyer JG, Urbanowicz RJ, Martin PCN, O'Connor K, Li R, Peng PC, et al. ChatGPT and large language models in academia: opportunities and challenges. BioData Min 2023;16:1–11. https://doi.org/10.1186/S13040-023-00339-9/METRICS.

[21] Salazar GZ, Zúñiga D, Vindel CL, Yoong AM, Hincapie S, Zúñiga AB, et al. Efficacy of AI chats to determine an emergency: a comparison between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. Cureus 2023;15. https://doi.org/10.7759/CUREUS.45473.

[22] Arocena F, Sansone S, Alvarez N. Technological disruption and democracy in the twenty-first century. Eur J Futur Res 2022;10:1–11. https://doi.org/10.1186/S40309-022-00189-4/METRICS.

[23] Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23:e17187.

[24] Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Https://DoiOrg/101146/Annurev-Publhealth-040119-094127 2020;41:433–51. https://doi.org/10.1146/ANNUREV-PUBLHEALTH-040119-094127.

[25] Do Nascimento IJB, Pizarro AB, Almeida JM, Azzopardi-Muscat N, Gonçalves MA, Björklund M, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022;100:544. https://doi.org/10.2471/BLT.21.287654.

[26] Peng C, Yang X, Chen A, Smith KE, PourNejatian N, Costa AB, et al. A study of generative large language model for medical research and healthcare. Npj Digit Med 2023;6:1–10. https://doi.org/10.1038/s41746-023-00958-w.

[27] Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183:589–96. https://doi.org/10.1001/JAMAINTERNMED.2023.1838.

[28] Wei K, Fritz C, Rajasekaran K. Answering head and neck cancer questions: an assessment of ChatGPT responses. Am J Otolaryngol 2024;45:104085. https://doi.org/10.1016/J.AMJOTO.2023.104085.

[29] Daungsupawong H, Wiwanitkit V. ChatGPT's responses to questions related to epilepsy. Seizure Eur J Epilepsy 2024;114:105. https://doi.org/10.1016/j.seizure.2023.12.004.

[30] Rocha-Silva R, de Lima BE, José G, Cordeiro DF, Viana RB, Andrade MS, et al. The potential of large language model chatbots for application to epilepsy: let's talk about physical exercise. Epilepsy Behav Reports 2024;27:100692. https://doi.org/10.1016/J.EBR.2024.100692.

[31] Rocha-Silva R, Rodrigues MAM, Viana RB, Nakamoto FP, Vancini RL, Andrade MS, et al. Critical analysis of information provided by ChatGPT on lactate, exercise, fatigue, and muscle pain: current insights and future prospects for enhancement. Https://DoiOrg/101152/Advan000732024 2024. doi: 10.1152/ADVAN.00073.2024.

[32] Pimentel J, Tojal R, Morgado J. Epilepsy and physical exercise. Seizure 2015;25: 87–94. https://doi.org/10.1016/J.SEIZURE.2014.09.015.

[33] Keyword Tool Google Keyword Planner - Search FREE n.d. https://keywordtool.io/ (accessed December 11, 2023).

[34] Seizure aftercare: What to do, foods, rest, mental well-being & more. Med News Today 2023. https://www.medicalnewstoday.com/articles/what-to-do-after-a-seizure-to-feel-better-2#summary (accessed December 11, 2023).

[35] Staying Safe With Exercise and Sports | Epilepsy Foundation. Epilepsy Found n.d. https://www.epilepsy.com/preparedness-safety/staying-safe/safety-exercise-and-sports (accessed December 10, 2023).

[36] Frequently Asked Questions About Epilepsy | CDC. Centers Dis Control Prev n.d. https://www.cdc.gov/epilepsy/about/faq.htm#What is epilepsy? What is a seizure? (accessed December 8, 2023).

[37] ILAE. Welcome to the International League Against Epilepsy n.d. https://www.ilae.org/ (accessed December 8, 2023).

[38] A New Hope for Patients With Epilepsy. Yale Med 2021. https://www.yalemedicine.org/news/epilepsy-deep-brain-stimulation.

[39] Chirigati F. Co-occurrence and similarity revisited. Nat Comput Sci 2022;2:67. https://doi.org/10.1038/s43588-022-00206-3.

[40] Marchand P, Ratinaud P. L'analyse de similitude appliqueé aux corpus textueles: les primaires socialistes pour l'election présiden- 518 Camargo, B. V., Justo, A. M. tielle française. In Actes des 11eme Journées internationales d'Analyse statistique des Données Textuelles. Actes des 11eme Journées Int. d'Analyse Stat. des Données Textuelles., Liège, Belgique: 2012, p. 687–99.

[41] Ignatow G, Mihalcea R. Text mining: a guidebook for the social sciences. Text Min A Guideb Soc Sci 2017. https://doi.org/10.4135/9781483399782.

[42] Flesch R. A new readability yardstick. J Appl Psychol 1948;32:221–33. https://doi.org/10.1037/H0057532.

[43] Flesch Kincaid Calculator - Flesch Reading Ease Calculator n.d. https://charactercalculator.com/flesch-reading-ease/ (accessed February 26, 2024).

[44] Lahlou S. Text mining methods: an answer to Chartier and Meunier. Pap Soc Represent 2012;20.

[45] Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. J Exp Psychol Gen 2012;141:2–18. https://doi.org/10.1037/A0024338.

[46] Espírito-Santo H, Daniel F, Superior Miguel Torga I. Calcular e apresentar tamanhos do efeito em trabalhos científicos (3): Guia para reportar os tamanhos do efeito para análises de regressão e ANOVAs. Rev Port Investig Comport e Soc 2018; 4:43–60. doi: 10.31211/rpics.2018.4.1.72.

[47] Field A. Descobrindo a Estatística Usando o SPSS. 5a edição. Penso; 2020.

[48] Haukoos JS, Lewis RJ. Advanced statistics: bootstrapping confidence intervals for statistics with "difficult" distributions. Acad Emerg Med 2005;12:360–5. https://doi.org/10.1197/J.AEM.2004.11.018.

[49] WHO. Epilepsy. World Heal Organ 2024. https://www.who.int/news-room/fact-sheets/detail/epilepsy (accessed April 3, 2024).

[50] Escoffery C, DiIorio C, Yeager KA, McCarty F, Robinson E, Reisinger E, et al. Use of computers and the Internet for health information by patients with epilepsy. Epilepsy Behav 2008;12:109–14. https://doi.org/10.1016/J.YEBEH.2007.07.013.

[51] Roumeliotis KI, Tselikas ND. ChatGPT and Open-AI Models: a preliminary review. Futur Internet 2023;15:192. https://doi.org/10.3390/FI15060192.

[52] Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2019;41:433–51. https://doi.org/10.1146/ANNUREV-PUBLHEALTH-040119-094127/CITE/REFWORKS.

[53] Crossley SA, Balyan R, Liu J, Karter AJ, McNamara D, Schillinger D. Predicting the readability of physicians' secure messages toimprove health communication using novel linguistic features: findings from the ECLIPPSE study. J Commun Healthc 2020;13:1. https://doi.org/10.1080/17538068.2020.1822726.

[54] Epilepsy n.d. https://www.who.int/news-room/fact-sheets/detail/epilepsy (accessed March 22, 2024).

[55] Min A, Miller WR, Rocha LM, Börner K, Brattig Correia R, Shih PC. Understanding contexts and challenges of information management for epilepsy care. Proc SIGCHI Conf Hum Factors Comput Syst CHI Conf 2023;2023. doi: 10.1145/3544548.3580949.

[56] Kim H-W, Shin D-H, Kim J, Lee G-H, Cho JW. Assessing the performance of ChatGPT's responses to questions related to epilepsy: a cross-sectional study on natural language processing and medical information retrieval. Seizure - Eur J Epilepsy 2024;114:1–8. https://doi.org/10.1016/J.SEIZURE.2023.11.013.

[57] Arida RM. Physical exercise and seizure activity. Biochim Biophys Acta - Mol Basis Dis 2021;1867:165979. https://doi.org/10.1016/J.BBADIS.2020.165979.

[58] Nakken KO, Bjørholt PG, Johannessen SI, LoSyning T, Lind E. Effect of physical training on aerobic capacity, seizure occurrence, and serum level of antiepileptic drugs in adults with epilepsy. Epilepsia 1990;31:88–94. https://doi.org/10.1111/J.1528-1157.1990.TB05365.X.

[59] Scheffer IE, Berkovic S, Capovilla G, Connolly MB, French J, Guilhoto L, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. Epilepsia 2017;58:512–21. https://doi.org/10.1111/EPI.13709.

[60] Wang X, Sanders HM, Liu Y, Seang K, Tran BX, Atanasov AG, et al. ChatGPT: promise and challenges for deployment in low- and middle-income countries. Lancet Reg Heal West Pacific 2023;41:100905. https://doi.org/10.1016/J.LANWPC.2023.100905.