

Çocuk Dergisi Journal of Child

Submitted: 14.02.2025
Revision Requested: 07.04.2025
Last Revision Received: 18.04.2025
Accepted: 25.04.2025

Research Article

Open Access

Evaluating Artificial Intelligence Responses to Questions About Febrile Seizures from a Family Perspective



Orhan Coskun¹  , Öznur Özaltın² , Berker Okay³  & Zübeyir Kavcar⁴ 

¹ Health Sciences University Gaziosmanpaşa Training and Research Hospital, Pediatric Neurology, İstanbul, Türkiye

² Hacettepe University, Faculty of Science, Department of Actuarial Science, Ankara, Türkiye

³ Medipol University Pediatrics, İstanbul, Türkiye

⁴ Health Sciences University Gaziosmanpaşa Training and Research Hospital Pediatrics, İstanbul, Türkiye

Abstract

Objective: Febrile seizures constitute the most frequently encountered seizure type in childhood, often leading families to seek information and clarification.

Materials and Methods: This study focuses on analyzing the content generated by the ChatGPT platform (<https://chat.openai.com/>) in response to queries related to febrile seizures.

Results: ChatGPT's answers to the questions about patient information were generally found to be satisfactory. Some questions received lower scores from the participating physicians due to confusing expressions or brevity of the answers. However, overall, it was deemed sufficient for patient information.

Conclusion: Providing accurate information to patients through artificial intelligence (AI) in common diseases may be beneficial for both patients and physicians. Although its reliability is currently subject to debate, it can improve over time.

Keywords

Febrile seizure · ChatGPT · Artificial intelligence · Parent information



“ Citation: Coskun, O., Özaltın, Ö., Okay, B. & Kavcar, Z. Evaluating Artificial Intelligence Responses to Questions About Febrile Seizures from a Family Perspective. Çocuk Dergisi–Journal of Child 2025; 25(2): 76–81. DOI: 10.26650/jchild.2025.1639778

© This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License. 

© 2025. Coskun, O., Özaltın, Ö., Okay, B. & Kavcar, Z.

✉ Corresponding author: Orhan Coskun dr.orhancoskun@hotmail.com



INTRODUCTION

Febrile seizures (FS) represent the most prevalent type of seizures during early childhood. These episodes are triggered by a fever exceeding 38°C and are not associated with infections of the central nervous system, electrolyte disturbances, or head trauma. FS typically occurs between the ages of 6 months and 5 years, with the highest incidence observed around 18 months. Epidemiological data indicate a higher occurrence in boys than in girls (1,2).

Key risk factors associated with febrile seizures (FS) include viral infections, certain vaccinations, hereditary susceptibility, prenatal exposure to maternal smoking, elevated maternal stress levels, and prolonged hospitalization in a neonatal intensive care unit exceeding 28 days (3,4).

Since the threshold level varies for each child, the highest fever value that can cause FS also varies. Although the degree of fever is the most important factor, seizures typically occur as the child's temperature rises. FS may be the first sign that a child is ill, and shortly after the seizure, the fever may be above 38 degrees (5,6).

Artificial intelligence (AI) has indeed rapidly transformed various domains, and medicine is no exception. Its impact ranges from improving diagnostics and treatment to enhancing patient care and administrative tasks. The capacity of AI to process extensive datasets, detect patterns, and generate predictions has significantly transformed various fields within healthcare. AI may be defined as technologies or machines capable of performing tasks like problem-solving and learning, language interpretation, pattern recognition, and planning, similar to human cognitive function (7). Recently, easily accessible AI applications such as ChatGPT have entered our lives. They have the capacity to generate human-like text, answer questions across various domains, and engage in natural language conversation. (8). Because FS are commonly observed, families frequently search the internet for information on this topic. In these situations where AI is also used, it is important to provide families with accurate and sufficient information. This study aimed to evaluate the adequacy of ChatGPT's responses to frequently asked questions about FS from the parents' perspective.

MATERIALS AND METHODS

First, we created a Google account that had never been used before. This account had no search history. Then, using this account, we determined the most frequently asked questions about FS by patients' families via a Google search. The questions were asked to ChatGPT (version 2.0 and 4.0) once on 29.11.2024. The initial responses provided were recorded.

To avoid potential bias from previous answers, each question was asked in a separate session at different times. This design ensured that earlier questions did not influence ChatGPT's responses to later ones.

The researchers recorded the answers received. These recorded answers were then turned into a survey. The questions and answers were shown to 30 pediatricians. They were asked if these answers were sufficient for their families. The responses were evaluated by 30 pediatricians using a 5-point Likert scale (Table 1). 1: very inadequate - 5: very adequate. Physicians rated the adequacy of each response from 1 (very inadequate) to 5 (more than adequate) based on accuracy, clarity, and clinical usefulness. The average scores were calculated. Statistical analyses were performed using IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Chicago IL, USA, 2015). Descriptive statistics were used to summarize the data, including the mean, standard deviation, and frequency. The Wilcoxon signed-rank test was applied to compare non-normally distributed paired data. A p-value of <0.05 was considered statistically significant.

Because our study does not contain any patient data and only involves the evaluation of ChatGPT responses by physicians, ethical approval was not required.

Table 1. ChatGPT's response evaluation by physicians (9)

Criteria	Definition
1. Very Inadequate	Significantly inaccurate; perhaps dangerous
2. Inadequate	Inaccurate or specific parts incoherent
3. Average	Modest inaccuracies or mistakes
4. Adequate	Sufficient but devoid of potentially useful details
5. More than adequate	Ensuring compliance with the information provided by pediatric neurologists and pediatricians.

RESULTS

The nine most frequently asked questions regarding FS were determined and are presented in Table 2. Physicians evaluated the adequacy of the answers provided by both the ChatGPT-2.0 and ChatGPT-4.0 versions based on these questions.

The distribution of physicians' evaluations for the ChatGPT-2.0 and ChatGPT-4.0 responses is shown in Tables 3 and 4, respectively. Overall, the ChatGPT-4.0 responses received higher adequacy scores compared to the ChatGPT-2.0.

Detailed examination of Table 2 reveals that the questions mainly reflect parents' concerns about FS and their desire to receive clear and prompt information regarding their child's condition. Although ChatGPT provides instant responses to

commonly asked questions, this study primarily assessed the reliability and accuracy of those responses. The answers generated by ChatGPT were evaluated by pediatricians, and the findings are summarized in Tables 3 and 4. The comparison of physicians' evaluations of the ChatGPT-2.0 and ChatGPT-4.0 responses is presented in Table 5. The Wilcoxon signed-rank test was used to assess differences in the adequacy scores for each of the nine frequently asked questions.

According to the Wilcoxon signed-rank test, there was no significant difference between the ChatGPT-2.0 and ChatGPT-4.0 responses for Question 1 ($z = -1.32$, $p =$

0.186). However, ChatGPT-4.0 responses were significantly rated higher than ChatGPT-2.0 for all other questions: Question 2 ($z = -3.63$, $p < 0.001$), Question 3 ($z = -4.35$, $p < 0.001$), Question 4 ($z = -4.53$, $p < 0.001$), Question 5 ($z = -4.63$, $p < 0.001$), Question 6 ($z = -3.05$, $p < 0.001$), Question 7 ($z = -4.43$, $p < 0.001$), Question 8 ($z = -4.77$, $p < 0.001$), and Question 9 ($z = -4.20$, $p < 0.001$).

Specifically, the median score for ChatGPT-4.0 responses was consistently 4 (adequate) across all questions, whereas ChatGPT-2.0 responses often received median scores of 2 or 3, reflecting perceptions of inadequacy or moderate adequacy.

Table 2. Most frequently asked questions about febrile seizures

Number of questions	Definition
Q1	What is the most likely cause of child's seizure?
Q2	What tests does my child need? Do these tests require special preparation?
Q3	Is this likely to happen again?
Q4	Does the child need treatment?
Q5	Will give a child fever-reducing medications during an illness help prevent febrile seizures?
Q6	What should I do the next time the child has a fever?
Q7	What can I do to help the child during a febrile seizure?
Q8	The child has another health condition. How can we manage them together?
Q9	Are there brochures or other printed materials I can take? What websites do you recommend?

Table 3. Pediatricians' evaluations of ChatGPT 2.0 responses

Number of questions	1. Very Inadequate	2. Inadequate	3. Average	4. Adequate	5. More than adequate
Q1 (percentage)	1 (3.3%)	5 (16.7%)	11 (36.7%)	13 (43.3%)	0 (0.0%)
Q2 (percentage)	0 (0.0%)	14 (46.7%)	10 (33.3%)	6 (20.0%)	0 (0.0%)
Q3 (percentage)	1 (3.3%)	8 (26.7%)	18 (60.0%)	3 (10.0%)	0 (0.0%)
Q4 (percentage)	4 (13.3%)	14 (46.7%)	9 (30.0%)	3 (10.0%)	0 (0.0%)
Q5 (percentage)	2 (6.7%)	7 (23.3%)	19 (63.3%)	2 (6.7%)	0 (0.0%)
Q6 (percentage)	0 (0.0%)	9 (30.0%)	13 (43.3%)	8 (26.7%)	0 (0.0%)
Q7 (percentage)	0 (0.0%)	5 (16.7%)	21 (70.0%)	4 (13.3%)	0 (0.0%)
Q8 (percentage)	0 (0.0%)	4 (13.3%)	23 (76.7%)	3 (10.0%)	0 (0.0%)
Q9 (percentage)	0 (0.0%)	5 (16.7%)	18 (60.0%)	7 (23.3%)	0 (0.0%)

Table 4. Pediatricians' evaluations of ChatGPT 4.0 responses

Number of questions	1. Very Inadequate	2. Inadequate	3. Average	4. Adequate	5. More than adequate
Q1 (percentage)	0 (0.0%)	3 (10.0%)	9 (30.0%)	18 (60.0%)	0 (0.0%)
Q2 (percentage)	0 (0.0%)	2 (6.7%)	7 (23.3%)	21 (70.0%)	0 (0.0%)
Q3 (percentage)	0 (0.0%)	1 (3.3%)	4 (13.3%)	25 (83.3%)	0 (0.0%)
Q4 (percentage)	0 (0.0%)	0 (0.0%)	5 (16.7%)	23 (76.7%)	2 (6.7%)
Q5 (percentage)	0 (0.0%)	1 (3.3%)	2 (6.7%)	26 (86.7%)	1 (3.3%)
Q6 (percentage)	0 (0.0%)	2 (6.7%)	8 (26.7%)	19 (63.3%)	1 (3.3%)
Q7 (percentage)	0 (0.0%)	0 (0.0%)	3 (10.0%)	21 (70.0%)	6 (20.0%)
Q8 (percentage)	0 (0.0%)	0 (0.0%)	3 (10.0%)	27 (90.0%)	0 (0.0%)
Q9 (percentage)	0 (0.0%)	0 (0.0%)	4 (13.3%)	18 (60.0%)	8 (26.7%)

Table 5. Comparison of Scores for Responses to Questions About Febrile Seizures in Children Between the ChatGPT-4.0 and ChatGPT-2.0 versions

	ChatGPT2.0				ChatGPT4.0			
	N	Mean	SD	Median	Mean	SD	Median	p ^a :
question1	30	3.2	0.85	3	3.5	0.68	4	0.186
question2	30	2.73	0.78	3	3.63	0.61	4	0.000
question3	30	2.77	0.68	3	3.8	0.48	4	0.000
question4	30	2.37	0.85	2	3.9	0.48	4	0.000
question5	30	2.7	0.7	3	3.9	0.48	4	0.000
question6	30	2.97	0.76	3	3.63	0.67	4	0.000
question7	30	2.97	0.56	3	4.1	0.55	4	0.000
question8	30	2.97	0.49	3	3.9	0.31	4	0.000
question9	30	3.07	0.64	3	4.13	0.63	4	0.000

N: Number of doctor evaluating the responses to the questions, SD: Standard Deviation

^aWilcoxon signed-rank test, significance $p < 0.005$

DISCUSSION

FS is the most common abnormal brain activity observed during the developmental process. While 50% of children with a history of FS experience their first seizure between 12 and 30 months, only 6-15% experience their first seizure after the age of 4 (10). The occurrence of FS differs across various global regions. In China, the rate has been reported as 0.35%, while in Finland it stands at 6.9%, and in India, it is observed at 10% (11).

Given the common occurrence of FS, especially in the pediatric population, it is likely that families may seek information about this condition from various sources, including AI-based platforms. Ensuring that parents have access to accurate and reliable information is essential in reducing anxiety and guiding appropriate management. Therefore, this study aimed to evaluate the adequacy and accuracy of the information provided by ChatGPT regarding FS.

The internet is one of the most important tools used by today's people to access information. We already know the role of "Dr. Google" in this matter. (12, 13).

Artificial Intelligence (AI) traditionally refers to the concept where computers can be trained to perform tasks through pattern recognition with minimal human intervention. A more contemporary definition of AI focuses on the use of algorithms that enable machines to tackle problems that once necessitated human cognitive abilities (14).

ChatGPT is an online chat robot that aims to respond to users' requests using artificial intelligence. ChatGPT has become a very popular and remarkable application, reaching more than millions of users in a very short time after its launch. (15). Additionally, as the data input of these chatbots increases, their capabilities improve and new versions are released. ChatGPT is capable of processing vast amounts of

data, identifying patterns in learning, generating content, retrieving relevant information, and translating it into the required language.

The integration of big data analytics into the healthcare sector offers substantial potential for enhancing clinical decision making, improving patient outcomes, and lowering healthcare costs. With the increasing digitization of health records and the widespread use of wearable technologies, the volume, variety, and velocity of healthcare data have grown dramatically. Big data technologies enable the processing of these complex datasets to uncover patterns, generate predictive insights, and support evidence-based medical practices. As a result, healthcare systems are shifting from reactive models of care toward more proactive and personalized approaches, driven by data-informed strategies for disease prevention and management (16).

Upon reviewing the physicians' evaluations of ChatGPT's responses in our study, a notable difference was observed between the answers provided by the ChatGPT 2.0 and 4.0 versions. In version 2.0, short and insufficient answers were noticeable. In version 4.0, however, the answers were more satisfying. In questions 1, 3, 4, and 5, the responses from ChatGPT 2.0 were found to be more inadequate by the participants. When asked for the reason, participants stated that the answers were insufficient and lacked detail. Additionally, participants particularly pointed out the absence of the information that FS can occur up to the age of 6.5 years, and that stating 5 years was not entirely correct. When asked why they preferred version 4.0 more, they mentioned that the answers were more detailed and contained important information regarding the differential diagnosis.

The reason the responses provided by version 2.0 were brief and occasionally insufficient is that earlier versions of the



model had more limited capacity. Version 2.0, particularly for more complex and detailed questions, generally provided brief and concise answers because the language model's capacity and ability to understand context were limited. As mentioned by the participants, the lack of specific information, particularly in important clinical areas like febrile convulsions, and the provision of incorrect or incomplete information reflects the more superficial answering style of the 2.0 model.

One of the main reasons for preferring version 4.0 is that its explanations are more detailed and, particularly, it provides critical information for differential diagnosis. For a pediatrician, it is not only important to have basic information about the illness but also to be able to make the correct differential diagnosis and accurately assess which symptoms point to which diseases. The emphasis that version 4.0 places on such important details has been seen as a major advantage by the participants.

The answers provided by ChatGPT 2.0 may not generally be sufficient for the family members of patients, especially since they are shorter and more superficial. Version 2.0 struggles to provide detailed answers to complex and in-depth questions, which can result in incomplete or unclear information, particularly in health-related matters. Family members typically require clear, understandable, and reliable information, and 2.0 might fall short in delivering such details. In health matters, family members often need more explanation, examples, and simplified information. For topics like FS and fever management, detailed, step-by-step instructions and advice are expected. 2.0's limitations in meeting these needs could lead to incorrect responses or insufficient information.

Version 4.0, on the other hand, with its more detailed answers and inclusion of clinically important information, would likely be more helpful for family members. This version provides more comprehensive and context-aware responses, making it more satisfying and guiding for someone seeking information.

In summary, the answers in version 2.0 may not be sufficient for family members, especially in situations that require more complex or detailed information.

Gilson et al. used ChatGPT in the USMLE and showed that artificial AI could answer questions at the level of almost 3rd year medical students. (17). In addition, many studies have been and are being conducted in different fields of medicine, based on the answers given by these chatbots to field-specific questions. (7,8,14). To date, no studies in the literature have assessed the responses of artificial intelligence regarding FS

from the perspective of the patient. In this regard, our study is the first of its kind in the literature.

However, it is currently a question mark whether such chatbots can define everything accurately, express it and convey it to the user. As the usage of ChatGPT increases and more data are uploaded, the responses to questions will be more accurate.

We as the whole world, are at a crossroads. human versus AI. We will see how artificial intelligence will gradually select the correct information over time.

When considering the limitations of the study, it should be noted that due to the constantly evolving and changing nature of AI, the quality of the responses may vary over time. Depending on how the questions are phrased, AI may provide more detailed answers. For this reason, we chose the most commonly asked question formats about FS found on Google. Comparisons with other AI programs can be made, and the responses provided by AI regarding various types of epilepsies can be evaluated in future studies.

CONCLUSION

Informing patients accurately with AI in common diseases will be beneficial for both patients and doctors. Although its reliability may be debatable at present, it will improve in the future. It has been observed that the answers given to the frequently asked questions about FS are on sufficient 4.0.



Ethics Committee Approval Ethical approval was not required for this study.

Peer Review Externally peer-reviewed.

Author Contributions Conception/Design of Study- O.C., Z.K.; Data Acquisition- Ö.Ö., B.O.; Data Analysis/ Interpretation- Ö.Ö., B.O.; Drafting Manuscript- O.C., Z.K, B.O.; Critical Revision of Manuscript- Ö.Ö.; Final Approval and Accountability- O.C., Z.K., B.O., Ö.Ö.

Conflict of Interest Authors declared no conflict of interest.

Financial Disclosure Authors declared no financial support.

Author Details

Orhan Coskun

¹ Health Sciences University Gaziosmanpaşa Training and Research Hospital, Pediatric Neurology, Istanbul, Türkiye

0000-0001-9229-404X dr.orhancoskun@hotmail.com

Öznur Özaltın

² Hacettepe University, Faculty of Science, Department of Actuarial Science, Ankara, Türkiye

0000-0001-9841-1702

Berker Okay


³ Medipol University Pediatrics, Istanbul, Türkiye

0000-0002-1506-9110



Zübeyir Kavcar

⁴ Health Sciences University Gaziosmanpaşa Training and Research Hospital
Pediatrics, Istanbul, Türkiye

 0000-0002-2476-4983

REFERENCES

- 1 Laino D, Mencaroni E, Esposito S. Management of pediatric febrile seizures. *Int J Environ Res Public Health*. 2018;15:2232. doi:10.3390/ijerph15102232
- 2 Sawires R, Buttery J, Fahey M.A Review of Febrile Seizures: Recent Advances in Understanding of Febrile Seizure Pathophysiology and Commonly Implicated Viral Triggers. *Front Pediatr*. 2021;9:801321.
- 3 Camfield P, Camfield C. Febrile seizures and genetic epilepsy with febrile seizures plus (GEFS+). *Epileptic Disord*. 2015 Jun;17(2):124-33.
- 4 Offringa M, Newton R, Cozijnsen MA, Nevitt SJ. Prophylactic drug management for febrile seizures in children. *Cochrane Database Syst Rev*. 2017 Feb 22;2(2): Update in Cochrane Database Syst Rev. 2021 Jun 16;6.
- 5 Subcommittee on Febrile Seizures; Febrile Seizures: Guideline for the Neurodiagnostic Evaluation of the Child With a Simple Febrile Seizure. *Pediatrics* February 2011; 127 (2): 389–394. 10.1542/peds.2010-3318.
- 6 Pavone P, Corsello G, Ruggieri M, Marino S, Marino S, Falsaperla R. Benign and severe early-life seizures: a round in the first year of life. *Ital J Pediatr*. 2018;44(1):54. doi:10.1186/s13052-018-0491-z
- 7 Nair PP, Aghoram R, Khilari ML. Applications of artificial intelligence in epilepsy. *Int J Adv Med Health Res*. 2021;8(2):41–8. doi:10.4103/ijamr.ijamr_94_21
- 8 Kim HW, Shin DH, Kim J, Lee GH, 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*. 2024;114:1–8.
- 9 Rogasch JM, Metzger G, Preisler M, Galler M, Thiele F, Brenner W, et al. ChatGPT: can you prepare my patients for [18F] FDG PET/CT and explain my reports? *J Nucl Med*. 2023;64(12):1876–9.
- 10 Tiwari A, Meshram RJ, Singh RK. Febrile seizures in children: a review. *Cureus*. 2022;14(11):e31509. doi:10.7759/cureus.31509.
- 11 Chin RF, Neville BG, Peckham C, Bedford H, Wade A, Scott RC; NLSTEPSS Collaborative Group. Incidence, cause, and short-term outcome of convulsive status epilepticus in childhood: a prospective population-based study. *Lancet*. 2006 Jul 15;368(9531):222–9.
- 12 Tan SS, Goonawardene N. Internet Health Information Seeking and the Patient-Physician Relationship: A Systematic Review. *J Med Internet Res*. 2017 Jan 19;19(1):e9.
- 13 Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C, A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *J Med Internet Res* 2013;15(4):e85.
- 14 Patel UK, Anwar A, Saleem S, Malik P, Rasul B, Patel K et al. Artificial intelligence as an emerging technology in the current care of neurological disorders. *J Neurol*. 2021;268(5):1623–42.
- 15 Cifarelli CP, Sheehan JP. Large language model artificial intelligence: the current state and future of ChatGPT in neuro-oncology publishing. *J Neurooncol*. 2023;163(2):473–4.
- 16 Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014 Feb 7;2:3.
- 17 Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. *JMIR Med Educ*. 2023;9:e45312.

