

The Evaluation of ChatGPT-4's Capacity to Provide Information on Febrile Seizures

Emine Özdemir Kaçer¹, Gürsel Şen²

¹Aksaray University Faculty of Medicine, Department of Pediatrics, Aksaray, Türkiye

²University of Health Sciences Türkiye, Evliya Çelebi Training and Research Hospital, Clinic of Pediatric Neurology, Kütahya, Türkiye

Abstract

Aim: This study aims to evaluate the capacity of ChatGPT-4 to provide accurate and reliable information about febrile seizures, focusing on its ability to deliver educational content and support caregivers and healthcare professionals.

Materials and Methods: A total of 30 questions related to febrile seizures were derived from the National Institute of Neurological Disorders and Stroke (NINDS) website. These questions were categorized into five main themes: "overview of febrile seizures," "symptoms and causes," "diagnosis," "treatment and management strategies," and "advice for families." Responses generated by ChatGPT-4 were assessed by experienced pediatricians and neurologists for accuracy and educational value, with results compared against ChatGPT-3.5 and the NINDS guidelines.

Results: Of the 30 responses evaluated, 22 were classified as "educationally valuable," 7 as "accurate but insufficient," and 1 as a "mix of correct and outdated information." None of the responses were deemed completely incorrect. In comparison with ChatGPT-3.5, ChatGPT-4 provided "better" responses for 8 questions, "similar" responses for 20, and "worse" responses for 2. Compared to the NINDS guidelines, ChatGPT-4 delivered comparable or superior responses in most cases, except for four questions where the official guidelines performed better due to richer context and graphical support.

Conclusion: ChatGPT-4 demonstrates substantial potential as an educational tool for febrile seizures, offering accurate and comprehensible information to caregivers and healthcare professionals. However, limitations such as the lack of detailed explanations and visual aids highlight the need for further development. Future research should explore broader datasets and user feedback to optimize these tools for personalized medical education.

Keywords: ChatGPT-4, febrile seizures, artificial intelligence in healthcare

Introduction

Febrile convulsions are defined as seizures with fever of 38°C or higher in children between the ages of 6 months and 5 years without central nervous system infection (1). The fact that it is seen in 2-5% of children makes this condition a common problem in pediatric neurology (2). Although it is generally considered benign, the recurrence rate during the same fever attack is 14.8% (3). Febrile convulsions are divided into two main groups: simple and complex. Simple febrile convulsions last less than 15 minutes, have a generalized seizure structure at the beginning, and occur only once in 24 hours. In contrast, complex febrile convulsions

may last longer than 15 minutes, have focal features, or recur more than once in 24 hours (4). Although febrile convulsions are generally considered self-limiting and "benign," they often prompt pediatric consultations. Management strategies may vary depending on the clinical context (5). High recurrence rates and unusual accompanying symptoms such as loss of consciousness or cyanosis increase the risk of traumatic accidents and the frequency of healthcare visits by increasing parents' anxiety levels. Therefore, providing accurate information and support mechanisms may contribute to more effective management of families during this process. Informing parents about the possible consequences of febrile convulsions and the development of self-



Corresponding Author: Emine Özdemir Kaçer MD, Aksaray University Faculty of Medicine, Department of Pediatrics, Aksaray, Türkiye
E-mail: dremineozdemir@hotmail.com **ORCID ID:** orcid.org/0000-0002-0111-1672

Cite this article as: Özdemir Kaçer E, Şen G. The Evaluation of ChatGPT-4's Capacity to Provide Information on Febrile Seizures. Eurasian J Emerg Med. 2025;24(1): 56-63.

Received: 04.01.2025

Accepted: 15.01.2025

Epub: 17.02.2025

Published: 19.03.2025



©Copyright 2025 The Emergency Physicians Association of Turkey / Eurasian Journal of Emergency Medicine published by Galenos Publishing House.
Licenced by Creative Commons Attribution-NonCommercial-NoDerivatives (CC BY-NC-ND) 4.0 International License.

management skills in daily life are critical. However, the gap between the information families need and the information they can access often leads them to alternative information sources such as web search engines (6,7). This lack of information can have potentially negative consequences such as misdirection and unnecessary anxiety.

Introduced in November 2022, ChatGPT is an artificial intelligence-based large-language model that can produce human-like responses to users' text-based inputs. Developed by OpenAI (OpenAI, L.L.C., San Francisco, CA, USA), this model is based on a generative pre-trained transformer (GPT) architecture, and is particularly notable for its capacity to provide text-based information flow (8). Its most recent version, GPT-4, was released in March 2023 and demonstrates superior accuracy in solving more complex problems due to its improved reasoning ability and broader knowledge base (9). ChatGPT has the potential to provide education and information in the field of healthcare. For example, it has been reported to be effective patient for patient education regarding conditions such as laryngopharyngeal reflux (10). However, the accuracy and reliability of knowledge-based responses remain a matter of debate, especially in clinical contexts. In common complaints, such as fever in children, ChatGPT responses have been stated to be high-quality, reliable, and understandable (11). However, studies on answering questions from patients with epilepsy have emphasized that the model provides both accurate and supportive answers but occasionally carries the risk of misleading users by providing incorrect information (12).

This study aimed to evaluate the accuracy and reliability of ChatGPT responses to questions frequently asked by parents and caregivers regarding febrile seizures. In addition, the performance of this model in providing educational information was analyzed, and its information capacity and problem-solving skills were also examined.

Materials and Methods

Data Source-Question Selection

For this study, questions were collected from the "febrile seizures" section of the National Institute of Neurological Disorders and Stroke (NINDS) website (13). This section provides a wide range of information covering symptoms, causes, risk factors, diagnosis, treatment, and management strategies related to febrile convulsions. A total of 30 questions were selected within the scope of the study and divided into the following categories: "overview of febrile convulsions," "symptoms and causes," "diagnosis," "treatment and management strategies," and "recommendations for families." Access to the NINDS

website was provided, where this content is publicly available (13). The selected content consisted of English texts and was translated into Turkish by a professional translator. During the translation process, texts were transferred verbatim preserving their accuracy and originality. These questions were based on the aim to support a better understanding of febrile convulsions to raise awareness among patients and families.

GPT Usage

The questions were integrated into the 4th version of ChatGPT on September 1, 2024. ChatGPT-4 is a language model developed by OpenAI that offers more advanced features than the previous versions. This version, which stands out for its improved reasoning ability, wider knowledge base, and superior performance in natural language processing tasks, can be accessed via OpenAI's official website or APIs. ChatGPT-4 offers a monthly subscription model and is used as a powerful tool for educational and research purposes.

The questions were prepared in English and grouped by category to analyze the relationships between them. Each question was entered into the system twice using the "New Chat" feature, and the answers were compared. In cases where no significant differences were observed between the answers, the initial answers were preferred for the analysis. In addition, the same questions were asked in ChatGPT version 3.5, and the responses were evaluated by comparing them with ChatGPT-4. This method was applied to analyze the performance differences between versions and examine the information-providing capacity of ChatGPT-4 more comprehensively.

Evaluation of Responses

The responses generated by a ChatGPT were reviewed by a pediatrician and a pediatric neurologist experienced in treating children with febrile seizures. Responses were evaluated for accuracy and educational value in four categories: "adequate and educationally valuable", "accurate but inadequate", "mix of correct/incorrect/outdated information", and "incorrect". In cases of disagreement between raters, a third reviewer with expertise in febrile seizure management was involved in making the final decision. In addition, ChatGPT-3.5 responses were compared with content from the NINDS "febrile seizures" section and with ChatGPT-4 responses. ChatGPT-4 performance was rated as "much better," "better," "similar," "worse," or "much worse" compared to other sources. Final evaluations were performed using a systematic approach, and expert opinions were used in cases of equivalence. This method was designed to analyze the information-providing capacity of ChatGPT-4 in detail and compare it with previous versions. This study was exempted from ethical review by the Institutional Review Board because it used

only publicly available programs and did not involve human participants. Patient consent was not obtained.

Statistical Analysis

All statistical analyses were conducted using SPSS version 22 (IBM SPSS Statistics for Windows, Armonk, NY: IBM Corp). Agreement between the two reviewers assessing ChatGPT's responses was calculated using the Weighted Cohen's Kappa coefficient. The Shapiro-Wilk test was used to assess the normality of the variables. Normally distributed data were reported as mean \pm standard deviation, while non-normally distributed data were presented as median (range). Categorical variables were expressed as frequencies and percentages. Pearson's correlation test was used for parametric data and Spearman rank correlation test for non-parametric data to evaluate relationships between variables.

Results

Evaluation of Responses from ChatGPT-4

The performance of the ChatGPT-4 was evaluated using 30 questions. Responses to 22 of these questions were classified as having "sufficient educational value," while 7 responses were found to be "correct but insufficient." Of the 22 responses that had 'sufficient educational value,' 21 were accepted by both raters and only one response had to be evaluated by a third expert. One response was evaluated as "a mixture of correct/incorrect/outdated information," but no response fell into the "incorrect" category (Table 1). When examined by category, all four responses under the heading "diagnosis" were evaluated as having "sufficient educational value". Three of the five responses in the "treatment and management strategies" category had "sufficient educational value," while two were considered "correct but insufficient." Four of the six responses under the heading "symptoms and causes" had "sufficient educational value," whereas two were included in the "correct but insufficient" category. Three of the five questions in the "febrile convulsions overview" category were rated as having "sufficient educational value," while the other two were found to be "correct but inadequate." All four responses under the "advice for families" heading were rated as having "sufficient educational value."

These findings suggest that ChatGPT-4 is effective in producing educational materials related to febrile convulsions. However, some categories were found to have gaps in the information and required more detail. Addressing these gaps could increase the model's educational capacity and strengthen its reliability in providing health information.

Comparison of ChatGPT-4 Responses with ChatGPT-3.5 Responses and Febrile Convulsions Guide

When the responses of ChatGPT-4 and ChatGPT-3.5 were compared in the evaluation of a total of 30 questions, no response from ChatGPT-4 was classified as "much better." However, eight responses were rated as "better," twenty responses were rated as "similar," and two responses were rated as "worse." These findings indicate that ChatGPT-4 generally performed similarly to or better than ChatGPT-3.5. In particular, ChatGPT-4 performed worse than ChatGPT-3.5 on the questions "impact of fever during febrile convulsion" and "Is long-term drug therapy appropriate for the management of febrile convulsions?" (Table 2). Compared to the official guideline, ChatGPT-4 provided a "much better" response to the question, "What should I do if my child has a febrile seizure?" In addition, it provided "better" responses to 16 questions and 'similar' responses to 10 questions. However, four responses were rated as 'worse' compared to the guideline, and no response fell into the "much worse" category. Questions on which official guidelines elaborated on included: "What is a febrile seizure?", "What is the risk of developing epilepsy in children with febrile seizures?", "How is long-term febrile seizure treatment planned?", "Should a child with febrile seizures go back to school?". Overall, ChatGPT-4 responses were found to be equivalent to or better than the official guidelines.

Discussion

To our knowledge, this is the first study to analyze the Chat GPT-4 responses related to febrile seizures. ChatGPT can interpret natural language inputs and produce responses that are appropriate for user needs. In this study, AI was evaluated using a dataset containing information and frequently asked questions about febrile convulsions. The results revealed that ChatGPT can provide accurate and comprehensive answers on topics such as the causes, symptoms, and management strategies of febrile convulsions. It also provides practical and understandable information about emergency measures to be taken during febrile convulsions and long-term treatment approaches.

In recent years, reliance on the Internet as a source of health information has increased. Individuals frequently use online resources to search for information about febrile convulsions, such as diagnoses, treatment options, and drug side effects (14).

In a study evaluating GPT-4, the model demonstrated higher accuracy rates compared to GPT-3.5 in the Japan Medical Licensing Examination, especially in general and clinical sentence questions. This suggests that GPT-4 can be an effective tool for medical education and clinical support in non-English-speaking countries such as Japan (15). In our study, the responses in the

Table 1. Performance evaluation of ChatGPT-4 responses					
Category	Questions	Sufficient educational value	Correct but inadequate	Mixed with correct/incorrect/outdated data	Completely incorrect
Febrile convulsions overview	What is febrile convulsion?	✓			
	What are the symptoms of febrile convulsion?	✓			
	Who is more likely to have febrile convulsions?	✓			
	Are febrile convulsions harmful?	✓			
	What is the risk of developing epilepsy in children who have febrile convulsions?		✓		
Symptoms and causes	What are the main causes of febrile convulsions?	✓			
	What changes occur in the body during febrile convulsions?	✓			
	Does fever trigger febrile convulsions?		✓		
	Is there a connection between febrile convulsions and family history?	✓			
	What are the typical symptoms of febrile convulsions?	✓			
Diagnosis	How is febrile convulsion diagnosed?	✓			
	What tests are used to diagnose febrile convulsions?	✓			
	Do brain imaging methods detect febrile convulsions?	✓			
	Are laboratory tests necessary to diagnose febrile convulsions?	✓			
Treatment and management strategies	What should be done during febrile convulsions?	✓			
	Should febrile convulsions be treated with medication?		✓		
	How is long-term febrile convulsion treatment planned?		✓		
	Does the use of anticonvulsant medications prevent febrile convulsions??	✓			
	What should be the treatment plan for children after febrile convulsions?	✓			
Advice for families	What should I do if my child has a febrile convulsion?	✓			
	What precautions can I take to prevent my child from having a febrile seizure?	✓			
	Should I see a doctor after a febrile convulsion?	✓			
	How are the daily lives of children with febrile convulsions affected?	✓			
	What do parents need to know to manage febrile convulsions?	✓			
	How should genetic predisposition within the family be assessed?	✓			
	How should children with febrile convulsions return to school?	✓			
	Does misinformation about febrile convulsions pose a risk to families?	✓			
	Do children need a special diet after a febrile convulsion?	✓			
	How are the psychological effects of febrile convulsions on the family managed?	✓			

Categories	Questions	ChatGPT-3.5 much better	ChatGPT-3.5 better	ChatGPT-3.5 similar	ChatGPT-3.5 worse	ChatGPT-3.5 much worse	Guidelines much better	Guidelines better	Guidelines similar	Guidelines worse
Overview of febrile seizures	What are febrile seizures?		✓	✓				✓	✓	
	What are the symptoms of febrile seizures?		✓	✓				✓	✓	
	Who is more likely to have febrile seizures?		✓	✓				✓	✓	
	Are febrile seizures harmful?		✓	✓				✓	✓	
	What is the risk of epilepsy in children with febrile seizures?			✓	✓			✓	✓	
Symptoms and causes	What are the main causes of febrile seizures?		✓	✓				✓	✓	
	What changes occur in the body during febrile seizures?		✓	✓				✓	✓	
	Does fever level trigger febrile seizures?			✓	✓			✓	✓	
	Is there a genetic link to febrile seizures?		✓	✓		✓		✓	✓	
	What are the typical symptoms of febrile seizures?		✓	✓				✓	✓	
Diagnosis	How are febrile seizures diagnosed?		✓	✓				✓	✓	
	What tests are used to diagnose febrile seizures?		✓	✓				✓	✓	
	Do brain imaging techniques detect febrile seizures?		✓	✓				✓	✓	
	Are lab tests necessary for diagnosing febrile seizures?			✓	✓			✓	✓	
	What should be done during a febrile seizure?		✓	✓				✓	✓	
Treatment and management	Should febrile seizures be treated with medication?			✓	✓			✓	✓	
	How should long-term treatment for febrile seizures be planned?			✓	✓			✓	✓	
	Does anticonvulsant medication prevent febrile seizures?		✓	✓				✓	✓	
	What is the treatment plan for children after febrile seizures?		✓	✓				✓	✓	

Table 2. Continued		Questions	ChatGPT-3.5 much better	ChatGPT-3.5 better	ChatGPT-3.5 similar	ChatGPT-3.5 worse	ChatGPT-3.5 much worse	Guidelines much better	Guidelines better	Guidelines similar	Guidelines worse
Advice for families		What should I do if my child has a febrile seizure?	✓	✓					✓	✓	
		What precautions can I take to prevent febrile seizures in my child?		✓	✓				✓	✓	
		Should I see a doctor after a febrile seizure?		✓	✓				✓	✓	
		How do febrile seizures affect daily life for children?		✓	✓				✓	✓	
		What should parents know about managing febrile seizures?		✓	✓				✓	✓	
		How should genetic predisposition in families be assessed?		✓	✓				✓	✓	
		When can children return to school after a febrile seizure?		✓	✓				✓	✓	
		Do misconceptions about febrile seizures pose risks for families?		✓	✓				✓	✓	
		Is a special diet needed for children after febrile seizures?		✓	✓				✓	✓	
		How should psychological impacts on families be managed?		✓	✓				✓	✓	

“general overview” category received a lower evaluation than the other categories. This was attributed to the fact that the information was generally presented in a listing format and in-depth explanations were lacking. For example, basic information was provided for the question “What is a febrile convulsion?” However, the details are insufficient. Despite this, clear listing of symptoms in questions such as “What are the symptoms of a febrile convulsion?” attracted attention, providing correct information. The genetic predisposition information was evaluated as “accurate but insufficient” since it was not sufficiently detailed in questions such as “Is there a connection between febrile convulsion and family history?” In general, ChatGPT-4 provided useful information in this category, but performed poorly against the visually supported explanations of the official guide.

In a study evaluating the electrocardiogram interpretation capabilities of emergency medicine specialists, cardiologists, and ChatGPT, GPT-4 was shown to be more successful than emergency medicine specialists in evaluating both daily and challenging ECG questions. It performed better than cardiologists in daily questions, but as the difficulty of the questions increased, its performance closely matched that of the cardiologists (16). In our study, the “diagnosis” category received a positive evaluation and all answers provided correct information. Although there is a risk of information obsolescence or inaccuracy when using artificial intelligence sources, it is noteworthy that it can provide accurate and useful information. A study evaluating GPT-4 to select antidepressant treatment for major depression reported that such models may create the risk of providing incorrect treatment options if used as a guide in sensitive areas such as psychopharmacological treatment without any human supervision or expert control (17). In our study, ChatGPT-4 provided generally useful and accurate information in the “treatment and management strategies”

category. Comprehensive answers to the questions “What should be done during a febrile convulsion?” and “What should be the treatment plan for children after febrile convulsions?” with sufficient educational value are needed. However, the answers to the questions “Should febrile convulsions be treated with medication?” and “How is long-term febrile convulsion treatment planned?” were evaluated as “correct, but insufficient” due to a lack of detail. The answers to the question “Does the use of anticonvulsant drugs prevent febrile convulsions?” were found to be detailed and comparable to the guidelines.

In a study evaluating the answers given by ChatGPT-4 to questions about epilepsy, it was emphasized that this model can be a valuable tool in conveying general medical information about epilepsy to the public. Although concerns about accuracy, copyright, and the ability of artificial intelligence to provide individual-specific information continue, it has been reported that such models can be used as auxiliary tools to reduce the educational burden of healthcare professionals in terms of patient education (6). In our study, in the “advice for families” category, ChatGPT-4 provided useful and accurate information on questions frequently asked by parents about febrile convulsions. Practical and understandable suggestions were provided for critical questions such as “What should I do if my child has a febrile convulsion?” and “What precautions can I take to prevent my child from having a febrile convulsion?” In addition, the answers given to the question “Should I see a doctor after a febrile convulsion?” were considered effective. The answers given to the question “How are the daily lives of children with febrile convulsions affected?” contain valuable information, both in terms of children’s adaptation to their daily lives and parents’ awareness. However, the official guide made the information easier to understand because of its graphic and visual support. The overall performance of ChatGPT-4 in this category was found to be satisfactory, but the lack of visual support led to a lower evaluation in some questions. Overall, this study demonstrates that generative AIs, such as ChatGPT, can be an effective tool for providing information and support for families and caregivers of children with febrile seizures. The increasing popularity of ChatGPT worldwide is expected to increase the demand for access to medical information through similar chatbot-based tools (18). As the performance of AI-powered chatbots improves and the amount of data used in their training processes increases, it will become increasingly critical to evaluate their ability to provide personalized responses appropriate to each patient’s situation.

Study Limitations

Although this study focused on evaluating the capacity of ChatGPT-4 to provide information about febrile convulsions, it had some limitations. First, the ChatGPT-4 responses were

evaluated based only on a selected dataset. Evaluations using larger datasets may provide more comprehensive information regarding the overall performance of the model. Second, the study was conducted only in English texts but was based on Turkish translations. The loss of meaning or changes in wording during the translation process may have affected the evaluation. The third limitation is that the responses were classified only by expert evaluation, and no evaluation was conducted from the user perspective. In addition, the accuracy of ChatGPT responses was examined only according to existing guidelines and expert knowledge. This does not completely eliminate the risks related to up-to-date information. Finally, factors such as lack of visual support may have limited the educational adequacy of the model.

Conclusion

This study, as one of the first to analyze the capacity of ChatGPT-4 to provide information on febrile convulsions, demonstrates that the model can provide accurate and comprehensive answers. ChatGPT-4, which has shown remarkable performance, especially in the categories of “treatment and management strategies” and “advice for families,” stands out as an effective tool in patient and family education. However, limitations such as a lack of detail and insufficient visual support in some questions indicate the need for improvement in information transfer.

AI-based chatbots have increasing potential to become important tools that facilitate access to information in healthcare services. For this potential to be fully realized, the capacity of models to provide personalized and error-free information appropriate for individuals needs improvement. Although ChatGPT-4, in its current form, is a supportive tool that can alleviate the educational burden of healthcare professionals, the effectiveness and usefulness of such models should be examined in greater depth with larger datasets and analyses based on user experience in future studies.

Ethics

Ethics Committee Approval: This study was exempted from ethical review by the Institutional Review Board because it used only publicly available programs and did not involve human participants. Patient consent was not obtained.

Informed Consent: Patient consent was not obtained.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.Ö.K., Concept: E.Ö.K., Design: E.Ö.K., G.Ş., Data Collection or Processing: E.Ö.K., Analysis or Interpretation: E.Ö.K., Literature Search: E.Ö.K., Writing: E.Ö.K.

Conflict of Interest: The authors declare that they have no conflict of interest.

Financial Disclosure: There are no financial conflicts of interest to disclose.

References

1. Subcommittee on febrile seizures; American Academy of Pediatrics. Neurodiagnostic evaluation of the child with a simple febrile seizure. *Pediatrics*. 2011;127:389-94.
2. Hauser WA. The prevalence and incidence of convulsive disorders in children. *Epilepsia*. 1994;35 Suppl 2:1-6.
3. Hirabayashi Y, Okumura A, Kondo T, Magota M, Kawabe S, Kando N, et al. Efficacy of a diazepam suppository at preventing febrile seizure recurrence during a single febrile illness. *Brain Dev*. 2009;31:414-8.
4. Practice parameter: the neurodiagnostic evaluation of the child with a first simple febrile seizure. American Academy of Pediatrics. Provisional Committee on Quality Improvement, Subcommittee on Febrile Seizures. *Pediatrics*. 1996;97:769-72; discussion 773-5.
5. Hampers LC, Trainor JL, Listernick R, Eddy JJ, Thompson DA, Sloan EP, et al. Setting-based practice variation in the management of simple febrile seizure. *Acad Emerg Med*. 2000;7:21-7.
6. 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.
7. Ozdemir Kacer E, Kacer I. Evaluating the quality and reliability of YouTube videos on scabies in children: a cross-sectional study. *PLoS One*. 2024 Oct 17;19:e0310508.
8. Sallam, M. ChatGPT utility in healthcare education, research, and practice: systematic review on the promising perspectives and valid concerns. in *Healthcare*. 2023.
9. openai.com. GPT-4. Available from: <https://openai.com/product/gpt-4>
10. Lechien JR, Carroll TL, Huston MN, Naunheim MR. ChatGPT-4 accuracy for patient education in laryngopharyngeal reflux. *Eur Arch Otorhinolaryngol*. 2024;281:2547-52.
11. Kacer EO, Ipekten F. Can ChatGPT provide quality information about fever in children? *J Paediatr Child Health*. 2025;61:60-5.
12. Wu Y, Zhang Z, Dong X, Hong S, Hu Y, Liang P, et al. Evaluating the performance of the language model ChatGPT in responding to common questions of people with epilepsy. *Epilepsy Behav*. 2024;151:109645.
13. Health, N.I.o. Available from: <https://www.ninds.nih.gov/health-information/disorders/febrile-seizures>
14. Oh J, You SY. Febrile seizure: what information can caregivers access through YouTube? *Seizure*. 202;91:91-6.
15. Takagi S, Watari T, Erabi A, Sakaguchi K. Performance of GPT-3.5 and GPT-4 on the Japanese Medical Licensing Examination: Comparison Study. *JMIR Med Educ*. 2023;9:e48002.
16. Günay S, Öztürk A, Özerol H, Yiğit Y, Erenler AK. Comparison of emergency medicine specialist, cardiologist, and chat-GPT in electrocardiography assessment. *Am J Emerg Med*. 2024;80:51-60.
17. Perlis RH. Research Letter: Application of GPT-4 to select next-step antidepressant treatment in major depression. *medRxiv* [Preprint]. 2023:2023.04.14.23288595.
18. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med*. 2023;388:1233-9.