

# Evaluating ChatGPT's effectiveness in providing medical information for pregnant women with rheumatic diseases

Romatizmal hastalığı olan hamile kadınlara tıbbi bilgi sağlamada ChatGPT'nin etkinliğinin değerlendirilmesi

<sup>1</sup>Bahar ÖZDEMİR ULUSOY<sup>1</sup>, <sup>2</sup>Can Ozan ULUSOY<sup>2</sup>

<sup>1</sup>Ankara Gaziler Physical Therapy and Rehabilitation Training and Research Hospital, Department of Rheumatology, Ankara, Türkiye

<sup>2</sup>Ankara Etlik City Hospital, Department of Perinatology, Ankara, Türkiye

## ABSTRACT

**Aim:** The growing use of ChatGPT as a source of health information highlights the need to assess its accuracy and adequacy. This study evaluated the accuracy and adequacy of ChatGPT (version 3.5) in responding to frequently asked questions from pregnant women with rheumatic diseases in both Turkish and English, aiming to assess its potential as a reliable source of patient information across languages in rheumatology and maternal-fetal medicine.

**Materials and Methods:** A total of 36 questions related to pregnancy and rheumatic diseases were obtained from Google and divided into seven subgroups. Questions were posed to ChatGPT in both Turkish and English and responses were evaluated on a 4-point scale by a rheumatologist (Expert 1) and a perinatologist (Expert 2). Mann-Whitney U test was used for statistical analysis ( $p < 0.05$  was considered significant).

**Results:** ChatGPT's English responses demonstrated a higher rate of accuracy and completeness compared to its Turkish responses. In English, 91.6% of answers were rated as correct, compared to 75.0% in Turkish. Expert 1 rated the average score for Turkish responses as  $3.64 \pm 0.54$  and for English responses as  $3.89 \pm 0.31$ , a difference that was statistically significant ( $p = 0.023$ ). Expert 2 rated Turkish responses with an average score of  $3.83 \pm 0.37$  and English responses with an average score of  $3.94 \pm 0.23$ , with no statistically significant difference ( $p = 0.136$ ).

**Conclusion:** ChatGPT demonstrates promise as an accessible source of information for pregnant women with rheumatic disease, but has limitations in its non-English responses. This highlights the need for improvement in language-specific training of language models. Further research is recommended to explore the performance of ChatGPT across multiple languages and medical specialties.

**Keywords:** ChatGPT, rheumatic diseases, pregnancy, language models, patient education

## ÖZ

**Amaç:** ChatGPT'nin bir sağlık bilgi kaynağı olarak artan kullanımı, doğruluğunun ve yeterliliğinin değerlendirilmesi ihtiyacını vurgulamaktadır. Bu çalışmada, ChatGPT'nin (versiyon 3.5) romatizmal hastalığı olan hamile kadınların sıkça sorduğu sorulara Türkçe ve İngilizce yanıt vermedeki doğruluğu ve yeterliliği değerlendirilerek, romatoloji ve anne-fetal tıbbi alanlarında farklı dillerde güvenilir bir hasta bilgi kaynağı olma potansiyeli değerlendirilmiştir.

**Gereç ve Yöntemler:** Gebelik ve romatizmal hastalıklarla ilgili toplam 36 soru Google'dan elde edildi ve yedi alt gruba ayrıldı. Sorular, ChatGPT'ye hem Türkçe hem de İngilizce olarak yöneltildi ve yanıtlar, bir romatolog (Uzman 1) ve bir perinatolog (Uzman 2) tarafından 4 puanlık bir ölçekle değerlendirildi. İstatistiksel analiz için Mann-Whitney U testi kullanıldı ( $p < 0.05$  anlamlı kabul edildi).

**Sonuçlar:** ChatGPT'nin İngilizce yanıtları, Türkçe yanıtlarına kıyasla daha yüksek bir doğruluk ve tamlik oranı göstermiştir. İngilizcede yanıtların %91,6'sı tam doğru olarak değerlendirilirken, Türkçede bu oran %75,0 olmuştur. Uzman 1, Türkçe yanıtlar için ortalama puanı  $3,64 \pm 0,54$  ve İngilizce yanıtlar için  $3,89 \pm 0,31$  olarak değerlendirmiştir; bu fark istatistiksel olarak anlamlıdır ( $p = 0,023$ ). Uzman 2, Türkçe yanıtları ortalama  $3,83 \pm 0,37$  ve İngilizce yanıtları ortalama  $3,94 \pm 0,23$  puanla değerlendirmiştir ve istatistiksel olarak anlamlı bir fark yoktur ( $p = 0,136$ ).

**Tartışma:** ChatGPT, romatizmal hastalığı olan hamile kadınlar için erişilebilir bir bilgi kaynağı olarak umut vaat etmekte, ancak İngilizce olmayan yanıtlarında sınırlamalar bulunmaktadır. Bu durum, dil modellerinin dile özgü eğitiminde iyileştirme gereğini vurgulamaktadır. ChatGPT'nin birden fazla dil ve tıbbi uzmanlık alanındaki performansını keşfetmek için daha fazla araştırma yapılması önerilmektedir.

**Anahtar Kelimeler:** ChatGPT, romatizmal hastalıklar, gebelik, dil modelleri, hasta eğitimi

**Cite as:** Özdemir Ulusoy B, Ulusoy CO. Evaluating ChatGPT's effectiveness in providing medical information for pregnant women with rheumatic diseases. Jinekoloji-Obstetrik ve Neonatoloji Tıp Dergisi 2025;22(1):38–44.

Geliş/Received: 01.11.2024 • Kabul/Accepted: 25.12.2024

Sorumlu Yazar/Corresponding Author: Bahar ÖZDEMİR ULUSOY, Ankara Gaziler Physical Therapy and Rehabilitation Training and Research Hospital, Department of Rheumatology, 06100 Ankara, Türkiye

E-mail: bharozdemir@gmail.com

Çevrimiçi Erişim/Available online at: <https://dergipark.org.tr/tr/pub/jgon>

## INTRODUCTION

In recent years, artificial intelligence (AI)-based large language models (LLMs) have revolutionized access to information and have begun providing guidance across various fields. One such model, ChatGPT, developed by OpenAI, is a conversational AI agent with powerful text-processing capabilities. Named the Generative Pretrained Transformer (GPT) 3.5, this model is designed to understand and respond to text-based questions, generate text, and perform various language-related tasks (1). ChatGPT can answer users' questions as if they were engaged in a conversation with a human. By synthesizing information from the internet, it presents complex topics in a summarized and understandable way, making it frequently used in many areas, including medical consultation (2).

One of ChatGPT's standout features is its ability to detect the language in which a question is asked and respond in the same language. This capability makes it an accessible and effective source of information for a global audience. Its clear and fluent writing style, combined with the ability to communicate in nearly any language, makes ChatGPT a versatile tool. In the medical field, ChatGPT is increasingly used for patient education and preliminary health guidance, offering accessible explanations for complex medical topics (3). For example, pregnant women with rheumatic diseases can turn to ChatGPT to learn about the effects of their condition on pregnancy and possible treatment options, receiving answers in their native language. However, the accuracy and medical adequacy of these responses require careful scrutiny to assess their reliability.

Pregnant women with rheumatic diseases frequently seek information on how their condition may affect pregnancy, medication safety, and potential risks during childbirth. To meet this need, many patients and their families turn to search engines or AI-based conversational agents. Especially during the sensitive period of pregnancy, receiving accurate responses to these inquiries is of great importance.

In this study, we identified the most frequently asked questions on Google by pregnant women with rheumatic diseases and posed these questions to the free version of ChatGPT in both Turkish and English. We then evaluated the responses on a 4-point scale based on our medical knowledge, current medical guidelines, and clinical experience as a rheumatologist and a perinatologist (4-9). Our evaluation focused on the scientific validity of ChatGPT's information, as well as the accuracy and depth with which it answered patients' questions. Our study aims to understand to what extent artificial intelligence can serve as a reliable source of information for patients and healthcare professionals in these specific medical fields.

## MATERIAL AND METHODS

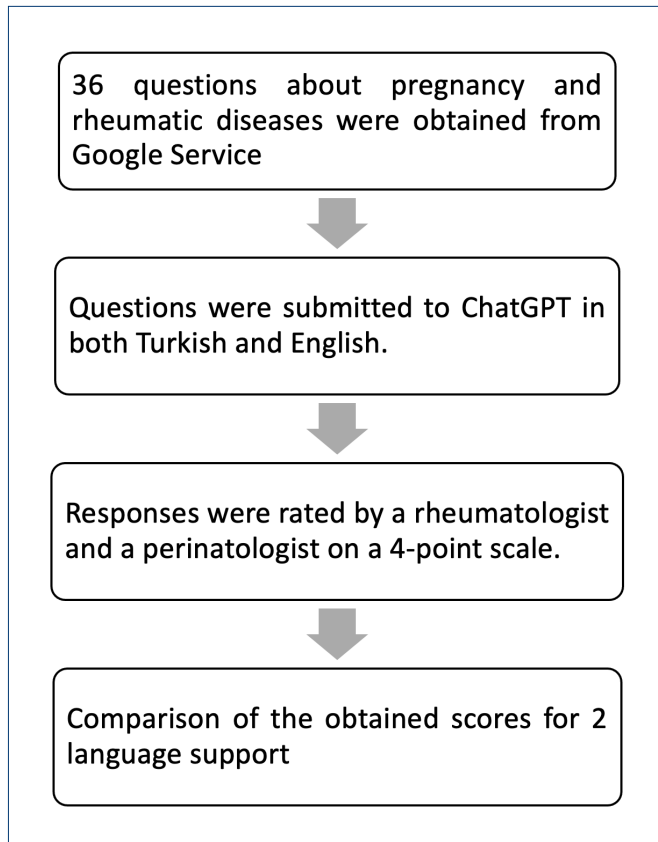
In this study, a total of 36 frequently asked questions regarding pregnancy and rheumatic diseases were obtained from Google services (10) and categorized into seven subcategories: "Basic Knowledge," "Ankylosing Spondylitis," "Rheumatoid Arthritis," "Psoriatic Arthritis," "Systemic Lupus Erythematosus,"

**Table 1.** List of questions asked by 2 experts to ChatGPT version 3.5

Question No	Question
1	Do rheumatic diseases affect my baby's development?
2	Is there a risk of premature birth due to rheumatic diseases?
3	Can rheumatic diseases cause congenital problems in my baby?
4	Can people with rheumatic diseases have children?
5	Is rheumatism medication used during pregnancy?
6	Does inflammation in the body harm the child?
7	Is high CRP an obstacle to getting pregnant?
8	How do rheumatic diseases affect my pregnancy?
9	Do rheumatic diseases worsen or improve during pregnancy?
10	Do rheumatic diseases require cesarean delivery?
11	Can a person with rheumatism have a normal delivery?
12	What should I do if my rheumatic disease flares up during pregnancy?
13	What can I do for rheumatic pains during pregnancy?
14	Is ankylosing spondylitis an obstacle to pregnancy?
15	Does ankylosing spondylitis cause problems during delivery?
16	Can ankylosing spondylitis be passed on to the baby?
17	Is rheumatoid arthritis exacerbated during pregnancy?
18	How does rheumatoid arthritis affect pregnancy?
19	Can rheumatoid arthritis be passed on to the baby?
20	Are the medicines used to treat rheumatoid arthritis safe during pregnancy?
21	Is psoriatic rheumatism an obstacle to pregnancy?
22	Does psoriatic rheumatism make it difficult to get pregnant?
23	Does psoriatic rheumatism worsen during pregnancy?
24	Can psoriatic rheumatism be passed on to the baby?
25	Does psoriatic rheumatism affect labor?
26	Are the medicines used to treat psoriatic rheumatism safe during pregnancy?
27	How does systemic lupus erythematosus (SLE) affect pregnancy?
28	Does lupus flare during pregnancy?
29	Can SLE be passed on to the baby?
30	Does scleroderma affect my ability to get pregnant?
31	What are the risks of scleroderma during pregnancy?
32	Can a mother with scleroderma pass the disease on to her baby?
33	Which medicines can a patient with scleroderma use during pregnancy?
34	Is colchicine used in pregnancy?
35	Does cortisone during pregnancy harm the baby?
36	Is biological therapy safe during pregnancy?

“Scleroderma,” and “Medication.” These questions, listed in Table 1, were directed to ChatGPT version 3.5 (OpenAI) in both Turkish and English (11). To maintain consistency and minimize variability, each question was asked once in both languages at the same time interval. This approach prevented potential variations in responses that might arise if the same question were asked multiple times or at different times, as ChatGPT can generate differing answers under such conditions. Responses were scored by two experts with at least 5 years of experience in their field: Expert 1, a rheumatologist, and Expert 2, a perinatologist. Answers were assessed on a 4-point scale based on completeness and alignment with established guidelines. A “4-point correct answer” was awarded for responses that were 100% complete and accurate according to medical guidelines. Answers with over 50% correct information were classified as “3-point partially correct answers,” while those with less than 50% accuracy received “2-point inadequate answer” scores. Responses containing any misinformation were rated as “1-point incorrect answers (Figure 1).”

Each expert independently reviewed and rated the responses in both languages, ensuring a consistent and objective evaluation process. The experts’ assessments reflected the completeness and reliability of the information provided by ChatGPT.



**Figure 1.** Flowchart of the planning of the study

## Statistical analyses

All statistical analyses were carried out using SPSS version 29 (SPSS Inc., Chicago, IL, USA). Statistical analyses included a normality test (Shapiro), and since the data did not follow a normal distribution, the Mann-Whitney U test was used. Statistical significance was defined as  $p < 0.05$ . This methodology ensured a rigorous evaluation of ChatGPT’s capacity to provide accurate and medically reliable information for pregnant individuals with rheumatic conditions.

## RESULTS

Table 2 shows the categorization of ChatGPT’s answers to the Turkish and English questions asked by Expert 1 and Expert 2 in terms of their accuracy. In the Basic Knowledge category, 100% of the English responses were rated correct by Expert 1, while Expert 2 found 92% of responses correct and 8% partially correct. Turkish responses, however, showed more variability; Expert 1 rated 54% as correct and 46% as partially correct, whereas Expert 2 found 92% correct and 8% partially correct.

In the Ankylosing Spondylitis category, all responses in both languages were rated as 100% correct by both experts. For the Rheumatoid Arthritis category, Expert 1 rated 75% of the English responses as correct and 25% as partially correct, while Expert 2 found all responses correct. In Turkish, both experts agreed, with 75% of responses rated as correct and 25% as partially correct.

For the Psoriatic Arthritis category, Expert 1 rated 83% of English responses as correct and 17% as partially correct, while Expert 2 found all responses correct. Turkish responses were rated 50% correct and 50% partially correct by Expert 1, while Expert 2 rated all responses as correct.

In the SLE category, both experts rated 100% of the English responses as correct. For Turkish responses, Expert 1 found 100% correct, whereas Expert 2 rated 75% correct and 25% partially correct. In the Scleroderma category, 75% of the English responses were rated correct by Expert 1, with 25% as partially correct, while Expert 2 found all responses correct. Turkish responses showed more variability; Expert 1 rated 75% as correct and 25% as partially correct, while Expert 2 rated 50% correct and 50% partially correct.

In the Medication category, both experts rated 67% of English responses as correct and 33% as partially correct. Turkish responses, however, displayed a more complex pattern: Expert 1 rated 67% as correct and 33% as mixed and misleading, while Expert 2 rated 67% correct and 33% partially correct.

**Table 2.** Evaluation of ChatGPT's responses to rheumatic disease questions during pregnancy in turkish and english by expert review

N (%)	English		Turkish	
N:36	Expert 1	Expert 2	Expert 1	Expert 2
<b>Basic Knowledge (n:13)</b>				
Partially correct	0 (0)	1 (8)	6 (46)	1 (8)
Correct	13 (100)	12 (92)	7 (54)	12 (92)
<b>Ankylosing Spondylitis (n:3)</b>				
Partially correct	0 (0)	0 (0)	0 (0)	0 (0)
Correct	3 (100)	3 (100)	3 (100)	3 (100)
<b>Rheumatoid Arthritis (n:4)</b>				
Partially correct	1 (25)	0 (0)	1 (25)	1 (25)
Correct	3 (75)	4 (100.0)	3 (75)	3 (75)
<b>Psoriatic Arthritis (n:6)</b>				
Partially correct	1 (17)	0 (0)	3 (50)	0 (0)
Correct	5 (83)	6 (100)	3 (50)	6 (100)
<b>Systemic Lupus Erythematosus (n:3)</b>				
Partially correct	0 (0)	0 (0)	0 (0)	1 (25)
Correct	3 (100)	3 (100)	3 (100)	3 (75)
<b>Scleroderma (n:4)</b>				
Partially correct	1 (25)	0 (0)	1 (25)	2 (50)
Correct	3 (75)	4 (100)	3 (75)	2 (50)
<b>Medication (n:3)</b>				
Mixed and misleading	0 (0)	0 (0)	1 (33)	0 (0)
Partially correct	1 (33)	1 (33)	0 (0)	1 (33)
Correct	2 (67)	2 (67)	2 (67)	2 (67)
<b>TOTAL</b>				
Mixed and misleading	0 (0.0%)	0 (0.0%)	1 (3.5%)	0 (0.0%)
Partially correct	4 (11.1%)	2 (5.6%)	11 (30.5%)	6 (16.6%)
Correct	32 (88.8%)	34 (94.4%)	24 (66.6%)	30 (83.3%)

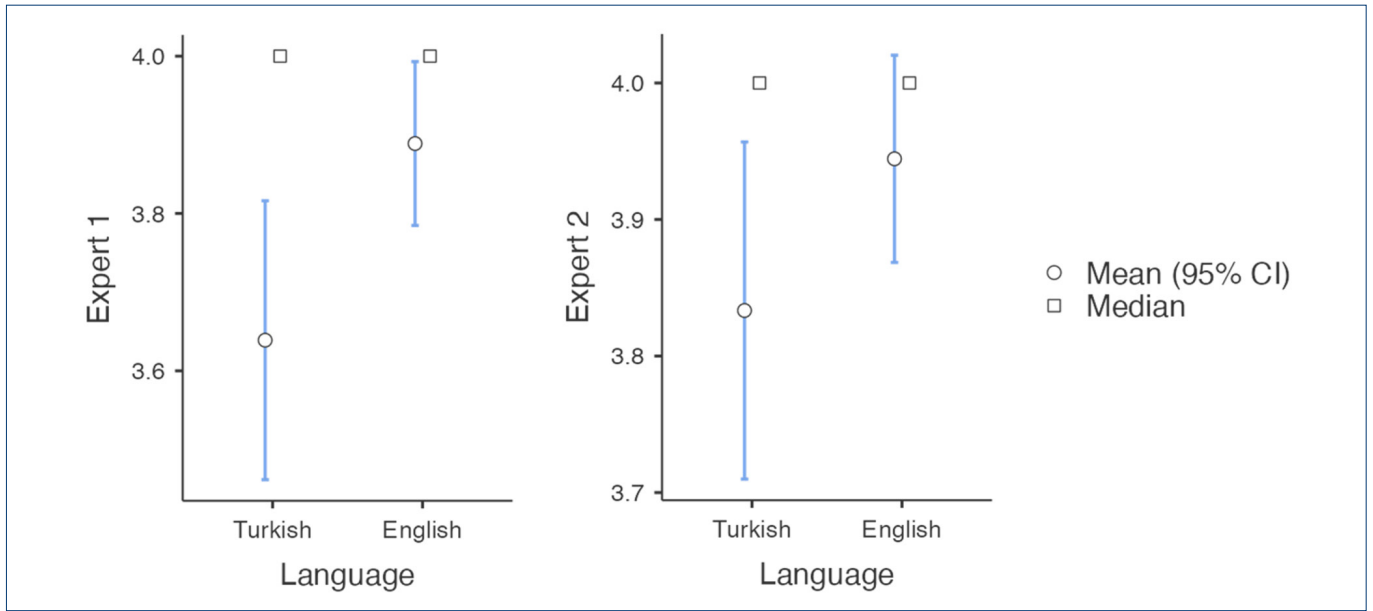
Expert 1: Rheumatologist; Expert 2: Perinatologist

**Table 3.** Comparison of mean scores for turkish and english responses by expert evaluation with standard deviation and statistical significance.

	Expert 1	Expert 2	P value
Turkish answers, mean $\pm$ SD	3.64 $\pm$ 0.54	3.83 $\pm$ 0.37	0.096
English answers, mean $\pm$ SD	3.89 $\pm$ 0.31	3.94 $\pm$ 0.23	0.397
P value	0.023	0.136	

In the total results, neither expert rated any responses as mixed and misleading in English (0%), whereas in Turkish, Expert 1 rated 3.5% as mixed and misleading. For partially correct responses, Expert 1 rated 11.1% in English and 30.5% in Turkish, while Expert 2 rated 5.6% in English and 16.6% in Turkish. Finally, correct responses were rated at 88.8% by Expert 1 and 94.4% by Expert 2 in English, compared to 66.6% by Expert 1 and 83.3% by Expert 2 in Turkish. ChatGPT answered 91.6% of the total 72 English questions asked by both experts correctly, while 8.4% were partially correct. In total 72 Turkish questions, 75.0% of them were correct, 23.6% were partial correct and 1.3% were mixed and misleading.

Table 3 shows the statistical analysis of the scores given by Expert 1 and Expert 2. Expert 1's average score for Turkish responses was  $3.64 \pm 0.54$ , compared to  $3.89 \pm 0.31$  for English responses, and this difference was statistically significant ( $p = 0.023$ ). Expert 2 rated Turkish responses with an average score of  $3.83 \pm 0.37$  and English responses with an average score of  $3.94 \pm 0.23$ , with no statistically significant difference ( $p = 0.136$ ). Additionally, there were notable differences between Expert 1 and Expert 2's scores within each language. For Turkish responses, Expert 2's average score ( $3.83 \pm 0.37$ ) was slightly higher than Expert 1's ( $3.64 \pm 0.54$ ), although this difference was not statistically significant ( $p =$



**Figure 2.** Mean and median scores with 95% confidence intervals for Turkish and English responses by expert evaluation

0.096). In the case of English responses, Expert 2 again awarded slightly higher average scores ( $3.94 \pm 0.23$ ) compared to Expert 1 ( $3.89 \pm 0.31$ ), but this difference also did not reach statistical significance ( $p = 0.397$ ).

Figure 2 illustrates the mean and median scores, along with the 95% confidence intervals, for ChatGPT's responses in both Turkish and English, as evaluated by Expert 1 and Expert 2. As shown, English responses consistently received slightly higher mean scores from both experts compared to Turkish responses. This aligns with the data presented in Table 3, where a statistically significant difference was observed between languages in Expert 1's evaluation ( $p = 0.023$ ), while Expert 2's evaluations did not show a significant difference ( $p = 0.136$ ).

## DISCUSSION

The findings of this study reveal that ChatGPT generally provides accurate or partially accurate responses to frequently asked questions regarding rheumatic diseases in pregnancy, both in English and Turkish. Although English responses demonstrate a higher rate of full accuracy, Turkish responses are also found to be satisfactory. However, notable differences in completeness and accuracy between the two languages suggest that ChatGPT's potential for providing information may vary depending on language support and the model's ability to grasp subtle linguistic nuances. These findings indicate that, while ChatGPT could be a valuable tool for patient education in healthcare, improvements in language support could enhance its effectiveness.

The use of LLMs like ChatGPT in the medical field is rapidly expanding, with ChatGPT frequently employed to respond to queries across diverse domains. Its application in medical contexts, including diagnosis, differential diagnosis, and interpretation of laboratory tests, has become widespread (12-14). Additionally, as examined in our study, it is also commonly utilized for patient counseling services. However, providing accurate and reliable information is crucial for these models, especially within the medical field. Moreover, as LLMs improve in accuracy, they offer advantages such as early disease detection, more precise differential diagnoses, and potential reductions in healthcare costs. Several studies in the literature assess ChatGPT's diagnostic evaluation capabilities.

In a study by Krusche et al., ChatGPT was compared with rheumatologists in differentiating inflammatory rheumatic diseases (IRDs) from other conditions. ChatGPT-4 was found capable of providing accurate differential diagnoses, achieving better sensitivity than a rheumatologist in identifying IRDs, underscoring its high potential as a tool for IRD differential diagnosis (15). Another study evaluated ChatGPT's accuracy and adequacy in answering rheumatology questions on a specialized medical entrance examination, demonstrating its value as a tool in rheumatology education with a 93.71% accuracy score (16). ChatGPT has set a new standard for both healthcare providers and patients seeking medical information. In one study, ChatGPT's ability to appropriately answer frequently asked questions about total hip arthroplasty was assessed, demonstrating its capability to provide evidence-based answers that were both effective and accessible to patients (17).



Consistent with numerous studies evaluating ChatGPT's performance in medical contexts, our study shows that ChatGPT displays high accuracy and adequacy in answering patients' most frequently asked questions. Parallel studies have highlighted limitations in ChatGPT's ability to provide detailed information on certain medical conditions. Carnino et al. evaluated ChatGPT's responses to questions from ear, nose, and throat patients in terms of accuracy, comprehensiveness, and bedside manner/empathy, pointing out limitations, especially in terms of accuracy and comprehensiveness (18). Another study demonstrated that ChatGPT falls short in managing special or highly specific cases, such as emergency urological cases (19). These results underscore the importance of thoroughly evaluating and critically assessing the information provided by ChatGPT.

ChatGPT's language comprehension and response abilities can vary depending on the language chosen, which may affect the accuracy and level of detail in its responses. Studies have examined how large language models are influenced by language selection, noting that reasoning and analytical abilities may be limited in certain languages (20). In a study by Yaslikaya and Kidi, ChatGPT's responses were evaluated in both Turkish and English for general information on breast cancer, treatment options, risks, and prevention. They observed that using English facilitated a clearer understanding of ChatGPT's answers (21). Similarly, in our study, we found that responses in English were more detailed and accurate than their Turkish counterparts (3.64 vs. 3.89 for Expert 1, 3.83 vs. 3.94 for Expert 2).

These findings indicate that ChatGPT's accuracy in addressing complex medical topics may diminish when used in languages other than English. A possible explanation for the variability in accuracy across languages lies in the underlying training data and model architecture. Since ChatGPT is primarily trained on English data, this may contribute to the higher accuracy observed in English responses compared to other languages. This suggests that, while ChatGPT may be a useful tool for preliminary information, caution is warranted when relying on non-English responses, especially in complex or sensitive medical fields such as rheumatology and maternal-fetal medicine.

The scoring discrepancies between Expert 1 (rheumatologist) and Expert 2 (perinatologist) also highlight the subjective nature of evaluating ChatGPT's responses. Although the scoring system provides a structured approach to assessing response quality, individual perspectives based on professional expertise inevitably influence the interpretation of adequacy and accuracy. Differences in scoring between the experts reflected their unique professional perspectives, with Expert 1 focusing on rheumatologic details and Expert 2 prioritizing aspects relevant to maternal-fetal health.

Our study offers important insights into the use of AI-based tools for patient education. Specifically, while ChatGPT's English responses in sensitive medical areas like pregnancy-related rheumatologic diseases are generally comprehensive, inconsistencies observed in its Turkish responses suggest that AI platforms may not yet fully replace human experts in providing reliable, linguistically sensitive medical guidance. This discrepancy raises concerns regarding health information equity, as patients who speak languages other than English may not receive the same level of accuracy or detail in AI-generated answers.

The limitations of our study should be acknowledged. We evaluated responses from ChatGPT's free version 3.5; therefore, results may vary in the paid or more advanced versions of the model. Additionally, while the questions were designed based on frequently asked questions related to rheumatic diseases in pregnancy, the phrasing of questions may vary significantly across users. Future research could investigate whether variations in question phrasing impact response accuracy or whether repeated questions lead to different answers. Despite efforts to maintain consistency, the subjective nature of scoring remains a factor influencing the interpretation of ChatGPT's performance.

In conclusion, while ChatGPT shows promise as a supplementary information source in the fields of rheumatology and maternal-fetal medicine, its limitations, particularly in languages other than English, should not be overlooked. Patients and healthcare professionals should proceed with an awareness of these potential shortcomings, ensuring that complex cases are verified by a healthcare provider. Further research is needed to assess the performance of future AI models across multiple languages and specialized medical domains.

#### Ethics declarations

This study did not involve human subjects and was therefore determined to be exempt from IRB review

#### Author Contributions

BOU and COU contributed to the conception and design of the study. BOU conducted the data collection and initial analysis. COU provided additional insights and critical revisions during the data analysis process. Both authors participated in the interpretation of results, drafting, and revising the manuscript. BOU prepared the initial draft, while COU reviewed and edited the final version. Both authors approved the final version of the manuscript for submission.

#### Acknowledgements

Not applicable

#### Disclosure statement

No potential conflict of interest was reported by the author(s).

#### Financial & Competing Interests Disclosure

There is no funding. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

## REFERENCES

1. <https://openai.com/index/chatgpt/> [Available from: <https://openai.com/index/chatgpt/>].
2. Brown TB. Language models are few-shot learners. arXiv preprint arXiv:2005.14165. 2020.
3. Patel SB, Lam K. ChatGPT: the future of discharge summaries? The Lancet Digital Health. 2023;5(3):e107-e8.
4. Sammaritano LR, Bermas BL, Chakravarty EE, Chambers C, Clowse ME, Lockshin MD, et al. 2020 American College of Rheumatology guideline for the management of reproductive health in rheumatic and musculoskeletal diseases. Arthritis & rheumatology. 2020;72(4):529-56.
5. Götestam Skorpen C, Hoeltzenbein M, Tincani A, Fischer-Betz R, Elefant E, Chambers C, et al. The EULAR points to consider for use of antirheumatic drugs before pregnancy, and during pregnancy and lactation. Ann Rheum Dis. 2016;75(5):795-810.
6. Peterson EA, Lynton J, Bernard A, Santillan MK, Bettendorf B. Rheumatologic Medication Use During Pregnancy. Obstet Gynecol. 2020;135(5):1161-76.
7. Rheumatology. 8th Edition ed September 29, 2022.
8. By Charles J. Lockwood M, MHCM Senior, Thomas Moore, MD, Joshua Copel, MD, Robert M Silver, MD, Robert Resnik, MD, Lorraine Dugoff, MD and Judette Louis, MD, MPH. Creasy-Resnik's Study Guide for Maternal Fetal Medicine. 9th Edition ed 04-10-2022.
9. Flint J, Panchal S, Hurrell A, van de Venne M, Gayed M, Schreiber K, et al. BSR and BHPR guideline on prescribing drugs in pregnancy and breastfeeding-Part I: standard and biologic disease modifying anti-rheumatic drugs and corticosteroids. Rheumatology (Oxford). 2016;55(9):1693-7.
10. <https://alsoasked.com> [Available from: <https://alsoasked.com>].
11. ChatGPT version 3.5 (OpenAI) [Available from: <https://chatgpt.com/g/g-FO0faAwkE-open-a-i-gpt-3-5>].
12. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56.
13. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. PLOS Digit Health. 2023;2(2):e0000198.
14. 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(13):1233-9.
15. Krusche M, Callhoff J, Knitza J, Ruffer N. Diagnostic accuracy of a large language model in rheumatology: comparison of physician and ChatGPT-4. Rheumatol Int. 2024;44(2):303-6.
16. Madrid-García A, Rosales-Rosado Z, Freitas-Nuñez D, Pérez-Sancristóbal I, Pato-Cour E, Plasencia-Rodríguez C, et al. Harnessing ChatGPT and GPT-4 for evaluating the rheumatology questions of the Spanish access exam to specialized medical training. Sci Rep. 2023;13(1):22129.
17. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT Responses to Common Patient Questions Regarding Total Hip Arthroplasty. J Bone Joint Surg Am. 2023;105(19):1519-26.
18. Carnino JM, Pellegrini WR, Willis M, Cohen MB, Paz-Lansberg M, Davis EM, et al. Assessing ChatGPT's Responses to Otolaryngology Patient Questions. Ann Otol Rhinol Laryngol. 2024;133(7):658-64.
19. Cocci A, Pezzoli M, Lo Re M, Russo GI, Asmundo MG, Fode M, et al. Quality of information and appropriateness of ChatGPT outputs for urology patients. Prostate Cancer Prostatic Dis. 2024;27(1):103-8.
20. Shi F, Suzgun M, Freitag M, Wang X, Srivats S, Vosoughi S, et al. Language models are multilingual chain-of-thought reasoners. arXiv preprint arXiv:2210.03057. 2022.
21. Yaslikaya S, Kidi MM. How much does ChatGPT know about breast cancer? Medicine Science. 2024;13(3).