



# Assessment and Cross-Language Comparison of ChatGPT's Responses to Frequently Asked Questions by Pregnant Women: A Multilingual Prospective Observational Study

*Gebelerin Sıkça Sorduğu Sorulara ChatGPT'nin Yanıtlarının Değerlendirilmesi ve Diller Arası Karşılaştırması: Çok Dilli Prospektif Gözlemsel Bir Çalışma*

Enes Serhat Coskun<sup>1</sup>, Ali Selcuk Yeniocak<sup>2</sup>, Havva Betül Bacak<sup>3</sup>, Erkan Aslan<sup>4</sup>, Fatma Ketenci Gencer<sup>3</sup>, Suleyman Salman<sup>3</sup>

<sup>1</sup>Department of Obstetrics and Gynecology, Biruni University Faculty of Medicine Hospital; <sup>2</sup>Department of Obstetrics and Gynecology, Basaksehir Cam And Sakura City Hospital; <sup>3</sup>Department of Obstetrics and Gynecology, Gaziosmanpasa Training and Research Hospital, Istanbul; <sup>4</sup>Department of Obstetrics and Gynecology, Buyuk Anadolu Meydan Hospital, Samsun, Türkiye

## ABSTRACT

**Aim:** To evaluate the efficacy of ChatGPT in providing pregnancy-related health information in English, Turkish, and Arabic, focusing on medical accuracy, empathetic communication, clarity, and appropriate referrals to obstetricians. This research aims to address concerns about variability in response accuracy and quality across languages.

**Materials and methods:** A prospective observational study was conducted within the Department of Obstetrics and Gynecology. Fifteen commonly asked pregnancy-related questions were translated into Turkish and Arabic by certified translators to ensure linguistic accuracy. Responses generated by ChatGPT in English, Turkish, and Arabic were independently assessed by four experienced obstetricians. Evaluation criteria included medical accuracy, empathetic tone, clarity, and appropriateness of referral. Statistical analysis was performed using chi-square tests, with statistical significance defined as  $p < 0.05$ .

**Results:** Statistically significant differences were observed across languages in medical accuracy ( $p=0.0012$ ), empathetic tone ( $p < 0.001$ ), and clarity ( $p < 0.001$ ). Responses provided in English were rated as "excellent" in 73.3% of cases, compared to 40% in Turkish and 41.6% in Arabic. Overall, more than 80% of all responses were rated as either "good" or "excellent." No significant difference was observed concerning the appropriateness of referral recommendations ( $p=0.0885$ ).

**Conclusion:** ChatGPT demonstrates potential as a reliable tool for addressing pregnancy-related health inquiries. However, discrepancies in response accuracy and clarity across languages highlight the need for improved language support, particularly for under-represented languages. Future research should include a broader range of questions and additional languages to enhance the reliability and clinical applicability of this technology.

**Key words:** artificial intelligence; artificial neural networks; ChatGPT; large language model; multilingual; pregnancy

## ÖZET

**Amaç:** Bu çalışmanın amacı, ChatGPT'nin gebelik ile ilgili sağlık bilgilerini İngilizce, Türkçe ve Arapça dillerinde sağlama konusundaki etkinliğini; tıbbi doğruluk, empatik iletişim, açıklık ve kadın doğum uzmanına yönlendirme uygunluğu açısından değerlendirmektir. Bu araştırma, farklı dillerde yanıtların doğruluğu ve kalitesiyle ilgili değişkenlik endişelerini ele almayı amaçlamaktadır.

**Gereç ve Yöntem:** Kadın Hastalıkları ve Doğum Anabilim Dalı bünyesinde prospektif gözlemsel bir çalışma yürütülmüştür. Sıkça sorulan 15 gebelik ile ilgili soru, dilsel doğruluğu sağlamak amacıyla yeminli tercümanlar tarafından Türkçe ve Arapçaya çevrilmiştir. ChatGPT tarafından İngilizce, Türkçe ve Arapça olarak verilen yanıtlar, deneyimli dört kadın doğum uzmanı tarafından bağımsız olarak değerlendirilmiştir. Değerlendirme kriterleri tıbbi doğruluk, empatik ton, açıklık ve yönlendirme uygunluğunu içermektedir. İstatistiksel analizler ki-kare testleri kullanılarak yapılmış, istatistiksel anlamlılık  $p < 0,05$  olarak kabul edilmiştir.

**Bulgular:** Tıbbi doğruluk ( $p=0,0012$ ), empatik ton ( $p < 0,001$ ) ve açıklık ( $p < 0,001$ ) açısından diller arasında istatistiksel olarak anlamlı farklılıklar tespit edilmiştir. İngilizce verilen yanıtlar vakaların %73,3'ünde "mükemmel" olarak değerlendirilirken, bu oran Türkçe için %40, Arapça için ise %41,6 olmuştur. Genel olarak, tüm yanıtların %80'inden fazlası "iyi" veya "mükemmel" olarak değerlendirilmiştir. Yönlendirme önerilerinin uygunluğu açısından anlamlı bir fark bulunmamıştır ( $p=0,0885$ ).

**Sonuç:** ChatGPT, gebelik ile ilgili sağlık sorularını yanıtlamada güvenilir bir araç olma potansiyeli göstermektedir. Ancak, diller arasında yanıtların doğruluğu ve açıklığı açısından gözlemlenen farklılıklar, özellikle az temsil edilen diller için dil desteğinin iyileştirilmesi gerektiğini ortaya koymaktadır. Gelecekteki araştırmalar, daha geniş bir soru yelpazesi ve daha fazla dili içerecek şekilde genişletilmeli, böylece bu teknolojinin güvenilirliği ve klinik uygulanabilirliği artırılmalıdır.

**Anahtar kelimeler:** yapay zekâ; yapay sinir ağları; ChatGPT; büyük dil modeli; çok dilli; gebelik

**İletişim/Contact:** Enes Serhat Coşkun, Department of Obstetrics and Gynecology, Biruni University Faculty of Medicine Hospital, Istanbul, Türkiye  
• Tel: 0536 643 38 76 • E-mail: enesserhatcoskun@gmail.com • **Geliş/Received:** 25.03.2025 • **Kabul/Accepted:** 21.04.2026

**ORCID:** Enes Serhat Coşkun: 0000-0002-6166-5458 • Ali Selcuk Yeniocak: 0000-0002-8149-6348 • Havva Betül Bacak: 0000-0001-8516-5493 • Erkan Aslan: 0000-0002-1300-1884 • Fatma Ketenci Gencer: 0000-0002-6076-2563 • Suleyman Salman: 0000-0001-7090-6105

## Introduction

Large Language Models (LLMs) and Artificial Neural Networks (ANNs) are important artificial intelligence (AI) models with broad applications in language processing and data analysis. ANNs are mathematical models inspired by the neural connections in the human brain that process data by learning from large datasets, recognizing patterns, and supporting decision-making (1,2). In contrast, LLMs have transformed Natural Language Processing (NLP) (3). Trained on vast textual datasets, LLMs learn language structure, meaning, and context, enabling them to perform tasks such as text generation, translation, and summarization. Both models use advanced architectures such as transformers and have complementary features that accelerate data analysis and optimize learning processes (4). While ANNs are generally used for diverse data types, LLMs provide deeper analysis of word relationships and context, particularly in language-based tasks (5). Together, these technologies enable broader, more efficient AI applications (6).

ChatGPT rapidly gained popularity, reaching 100 million users within a few months of its launch in November 2022 and reaching 200 million weekly active users by 2025 (7). This growth reflects its broad capacity for information provision and interaction. ChatGPT is an AI system with advanced capabilities for understanding and generating human language through the integration of LLMs and ANNs (8). These technologies enable ChatGPT to communicate in different languages and generate customized responses that account for cultural, grammatical, and contextual differences (7). After training on extensive textual data, ChatGPT learns relationships between words and sentences and generates logical responses appropriate to a given question or topic (9). As a transformer-based model, it uses an attention mechanism to interpret the context of each word in the input text (10,11). For example, when a word has multiple meanings across languages, it can select the meaning most appropriate to the user's context. This multilingual capability is promising for education, international collaboration, and cultural communication (12). However, careful design and supervision are necessary to reduce misunderstandings arising from differences in language and context (13).

In developed countries, expectant mothers commonly obtain information about pregnancy through prenatal education classes, medical consultations, and online

research (14). Artificial intelligence is becoming increasingly important because it can provide rapid responses to expectant mothers' questions whenever and wherever needed. AI-powered tools, particularly large language models such as ChatGPT, are notable for delivering instant information (15). In recent years, ChatGPT has attracted considerable attention because of its large user base and potential role in health information (16,17).

## Objectives

Our primary objective was to evaluate ChatGPT's responses to pregnancy-related questions in three different languages in terms of accuracy, clarity and comprehensibility, empathetic tone, and appropriate referral to obstetricians. Our secondary objective was to compare ChatGPT's responses across these languages and assess its potential usability as an information source for expectant mothers.

## Material and Methods

### Study Design

This prospective observational study compared the responses provided by ChatGPT-4o across different languages to the most frequently asked questions by pregnant women worldwide, accounting for ethnic differences (Table 1).

### Setting

This study compares responses from ChatGPT-4o to questions asked between October 2024 and January 2025.

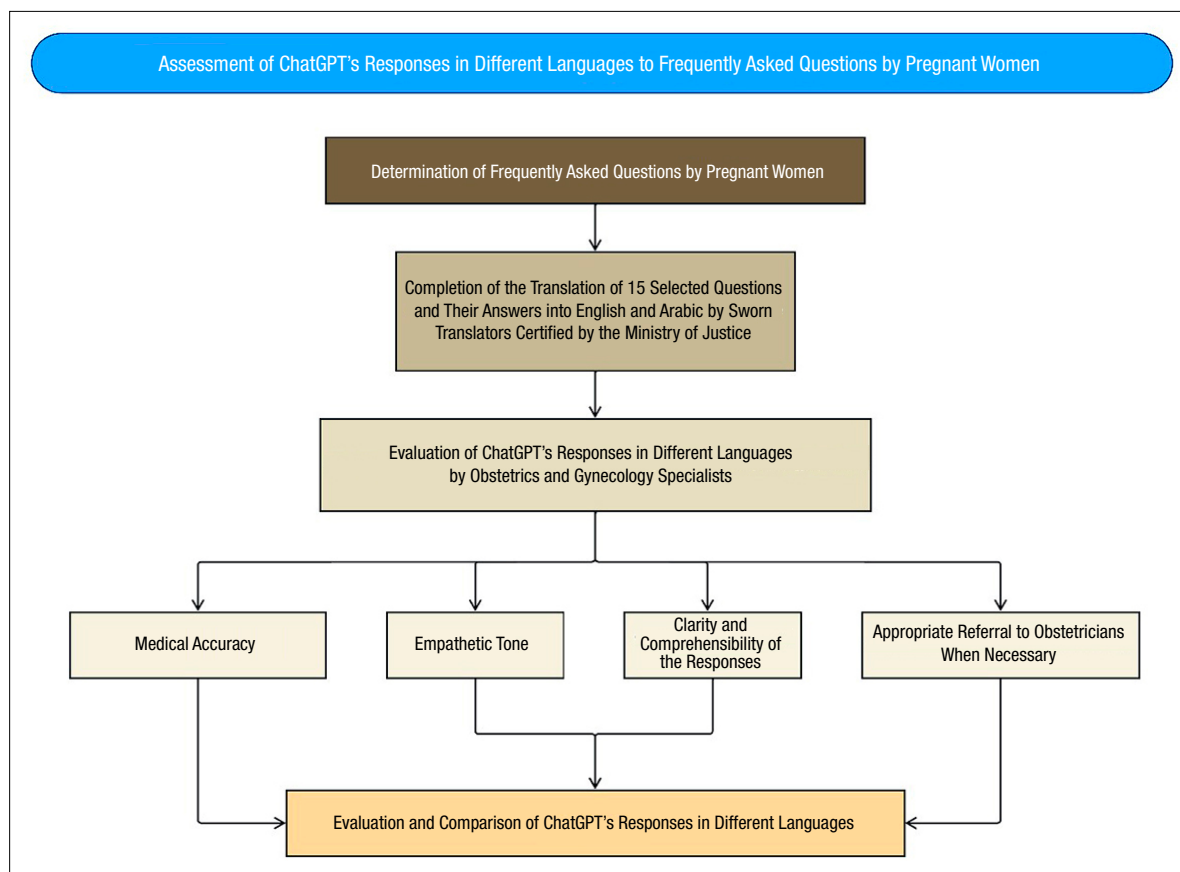
### Participants

To evaluate the usability of ChatGPT-4o in pregnancy education, responses were generated using an AI tool based on the ChatGPT-4o model, which is freely available and automatically accessible to users at initial launch. The study was conducted in accordance with the steps described below.

Frequently asked questions by expectant mothers during pregnancy were selected by four obstetricians based on their clinical experience. These questions were submitted to ChatGPT as if they were asked directly by pregnant patients. Common topics included pregnancy symptoms, prenatal care, nutrition, exercise recommendations, and types of delivery. All questions were translated into Turkish and Arabic by sworn

**Table 1.** Detailed comparison of questions and answers in three languages

| Questions                                                                                                                | Comments/ Observations                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|--------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Which food and drink should I consume or which should be avoided during pregnancy?                                       | In the English version, there was no mention of herbal teas, carbonated beverages, or desserts. While herbal teas were included in the Turkish version, no specific information regarding their dosage or quantity was provided. The Turkish content was further expanded by including references to probiotics and portion control. In contrast, the responses in the Arabic version were generally superficial and insufficient in detail.                                                                                                                                                                   |
| How do I know when labor has started? When should I go to the hospital?                                                  | The content provided in all three languages is generally sufficient; however, the English version includes practical information, such as essential items for a hospital bag, and specific guidelines, such as the “5-1-1 rule.” In the Turkish version, topics such as pelvic pressure and digestive changes were addressed. In contrast, the Arabic version remains more superficial in terms of the depth of information provided.                                                                                                                                                                          |
| Which is healthier? Should I have a spontaneous vaginal birth or a cesarean section?                                     | The English text provides more detailed, explanatory information, while the Turkish text focuses on more direct, practical guidance. Although the Arabic text adopts a more concise, summary-oriented approach, all versions have been deemed adequate. Additionally, the English version covers topics such as pelvic prolapse and incontinence.                                                                                                                                                                                                                                                              |
| What are the symptoms of postpartum depression?                                                                          | The English response includes postpartum depression risk factors, making it richer in content. In contrast, the Turkish and Arabic responses do not address risk factors or treatment options. The Arabic response appears to cover the topic more superficially.                                                                                                                                                                                                                                                                                                                                              |
| What screenings or tests should I have during pregnancy? Are they harmful to my baby?                                    | In the English and Turkish responses, detailed information is provided regarding the timing of tests, risk levels, and procedures, with specific mentions of NIPT, invasive CVS, and amniocentesis. However, such details are missing in the Arabic text.                                                                                                                                                                                                                                                                                                                                                      |
| Is it safe to have sexual intercourse during pregnancy? Does sexuality harm the baby?                                    | The Turkish response primarily focuses on listing complications and risks, while the Arabic response emphasizes the importance of emotional bonding and general safety. The English response adopts a more balanced approach, addressing both risks and safe practices comprehensively.                                                                                                                                                                                                                                                                                                                        |
| When can I start to feel my baby's movements during pregnancy?                                                           |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Is it normal to have nausea and back pain during pregnancy? When will it go away?                                        |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Should I take vitamins and supplements during pregnancy?                                                                 | The English response is more detailed, mentioning omega-3 sources and alternatives for vegans/vegetarians, which are missing in the Turkish and Arabic versions. It also includes dietary sources for supplements (e.g., folic acid from leafy greens), unlike the others. While the English version offers broader nutritional guidance, the Turkish and Arabic texts mainly focus on supplement use.                                                                                                                                                                                                         |
| Is it safe to exercise during pregnancy? Which exercise should I do? How long should the duration be?                    | In the Arabic and English responses, the duration of specific exercise types is emphasized more frequently, while the Turkish responses focus more on general principles and recommendations. Additionally, the English version addresses the risks of exercises performed while lying on the back and the issue of vena cava compression.                                                                                                                                                                                                                                                                     |
| Are the spots on my skin during pregnancy normal? Will these spots go away after birth?                                  | In the Arabic and English responses, skin blemishes and their possible appearances are explained in detail, whereas the Turkish response contains less detail and is insufficient in describing lesions.                                                                                                                                                                                                                                                                                                                                                                                                       |
| What is the recovery process like after normal vaginal birth and cesarean birth?                                         | In the English response, the distinction between CS and NSD is clearly made and explained separately. The Turkish response appears more complex, presenting the information without clearly differentiating the two. Although the Arabic response addresses CS and NSD individually, it lacks sufficient information regarding the recovery process and duration.                                                                                                                                                                                                                                              |
| Is it possible to relax during pregnancy with natural methods such as relaxation massage, yoga, and breathing exercises? | Although all three languages provide detailed, well-structured responses, the English version also covers “aromatherapy” and “warm water baths.” Despite its brevity, the Arabic response includes comprehensive and essential information.                                                                                                                                                                                                                                                                                                                                                                    |
| How much weight should I gain in pregnancy in kg?                                                                        | The Turkish response lacks information on special cases such as twin pregnancies and does not specify how the information is distributed across trimesters. In contrast, the Arabic response is more comprehensive and detailed compared to the other versions.                                                                                                                                                                                                                                                                                                                                                |
| Can I fast during pregnancy?                                                                                             | The English response primarily emphasizes the recommendation to avoid fasting. In contrast, the Arabic version examines fasting during each trimester individually and discusses the associated risks, including a statement suggesting that fasting may be safe in the third trimester if the patient is in good health. While all versions include the phrase “Prolonged fasting or dehydration can harm the health of the mother and baby,” the Turkish response offers alternative solutions and mentions religious flexibility, whereas the Arabic version lacks any mention of religious accommodations. |



**Figure 1.** This study consisted of four main phases. The first phase was "Determination of Frequently Asked Questions by Pregnant Women." The second phase involved "Completion of the Translation of 15 Selected Questions and Their Answers into English and Arabic by Sworn Translators Certified by the Ministry of Justice." The third phase was "Evaluation of ChatGPT's Responses in Different Languages by Obstetrics and Gynecology Specialists" based on key criteria, including medical accuracy, empathetic tone, clarity and comprehensibility of the responses, and appropriate referral to obstetricians when necessary. The fourth phase was "Evaluation and Comparison of ChatGPT's Responses in Different Languages."

translators approved by the Justice Commission in our country, and the translations were signed accordingly. The 15 most frequently asked pregnancy questions, in their original language versions, were directed to ChatGPT-4o in Turkish, English, and Arabic. The AI-generated responses were obtained in written form. These responses were then independently assessed in each language by four gynecologists and obstetricians, each with at least five years of experience. Turkish, Arabic, and English were selected because they are the three most commonly used languages in our country. When necessary, outpatient clinic services are provided in these languages with translator assistance (Fig. 1).

#### Data sources/Measurement

Responses generated by ChatGPT-4o were evaluated according to four main criteria: medical accuracy, empathetic tone, clarity and comprehensibility, and appropriate referral to obstetricians when necessary.

The responses in each language were independently assessed by four gynecologists and obstetricians, each with at least five years of experience. Each response was rated as "excellent," "good," "fair," or "poor," and evaluators were asked to comment on any deficiencies. The collected data were categorized according to these evaluation groups and organized into tables.

#### Ethical Considerations

This study evaluated only AI-generated responses, and no direct human data were used. The questions were formulated based on the previous clinical experiences of obstetrics and gynecology specialists. No patient information was provided to ChatGPT, and the questions were written in the first person to simulate how a patient would address a healthcare professional. All questions were anonymized, and no personal health data were processed; therefore, ethical review was deemed unnecessary. The required permission was

obtained from the Provincial Directorate of Health by the corresponding author on 06/12/24, with reference number E-43129869-514.99-261699120. The study was conducted in accordance with the Declaration of Helsinki.

### Statistical Methods

Data were analyzed using IBM Statistical Package for Social Sciences (SPSS) program version 15 for Mac-iOS. Descriptive statistics were calculated for categorical variables and presented as frequencies. Homogeneity was assessed using Levene's test. Pearson's chi-square test was used for categorical variables. A p-value <0.05 was considered statistically significant.

## Results

Responses to 15 pregnancy-related questions asked in three languages were compared. Four obstetricians evaluated the responses according to medical accuracy, empathetic tone, clarity and comprehensibility, and appropriate referral to obstetricians when necessary. Table 2 summarizes the evaluations across English, Turkish, and Arabic using these four criteria. Ratings were presented as percentages for each category ("excellent," "good," "fair," and "poor"), and chi-square tests were used to assess differences among languages. Significant differences were found in medical accuracy ( $p=0.0012$ ), empathetic tone ( $p<0.001$ ), and clarity and comprehensibility ( $p<0.001$ ). However, no significant difference was observed for appropriate referral to obstetricians ( $p=0.0885$ ) (Table 2).

For medical accuracy, English had a higher percentage of "excellent" responses than Turkish and Arabic, whereas Turkish and Arabic responses were more often rated as "good" or "fair." For empathetic tone, English again showed the highest proportion of "excellent" ratings, while Arabic had higher proportions of "good" and "fair" ratings. Regarding clarity and comprehensibility, all languages achieved high percentages of "excellent" ratings; however, Arabic also showed higher proportions of "good" and "fair" responses. These findings suggest that although ChatGPT-4o provides generally accurate responses in English, the quality and content of responses may vary across languages. For appropriate referral to obstetricians, no significant language-based difference was detected, indicating similar performance across all languages.

Overall, this study represents an important step in evaluating language-related variation in ChatGPT-4o responses. Future studies with more languages and larger question sets may help confirm the generalizability of these findings.

## Discussion

During pregnancy, accurate and timely answers to common questions are essential to help prevent high-risk situations and guide expectant mothers on when to seek obstetric care. Delayed responses to concerns about potential risks may result in serious and sometimes irreversible adverse outcomes for both mother and fetus. Therefore, technologies such as ChatGPT may serve as immediate sources of information for

**Table 2.** ChatGPT's evaluation of medical accuracy, empathy, clarity, and referrals by obstetric specialists

|                                                      | Language | Excellent | Good      | Fair     | Poor  | P value    |
|------------------------------------------------------|----------|-----------|-----------|----------|-------|------------|
| Medical accuracy                                     | English  | 44(%73.3) | 14(%23.3) | 2(%3.3)  | 0(%0) | $p=0.0012$ |
|                                                      | Turkish  | 24(%40)   | 29(%48.3) | 7(%11.6) | 0(%0) |            |
|                                                      | Arabic   | 25(%41.6) | 26(%43.3) | 9(%15)   | 0(%0) |            |
| Empathetic tone                                      | English  | 59(%98.3) | 1(%1.6)   | 0(%0)    | 0(%0) | $p<0.001$  |
|                                                      | Turkish  | 51(%85)   | 9(%15)    | 0(%0)    | 0(%0) |            |
|                                                      | Arabic   | 38(%63.3) | 18(%30)   | 4(%6)    | 0(%0) |            |
| Clarity and comprehensibility                        | English  | 57(%95)   | 3(%5)     | 0(%0)    | 0(%0) | $p<0.001$  |
|                                                      | Turkish  | 58(%96.6) | 2(%3.3)   | 0(%0)    | 0(%0) |            |
|                                                      | Arabic   | 44(%73.3) | 12(%20)   | 4(%6.6)  | 0(%0) |            |
| Appropriate referral to obstetricians when necessary | English  | 57(%95)   | 3(%5)     | 0(%0)    | 0(%0) | $p=0.0885$ |
|                                                      | Turkish  | 54(%90)   | 6(%10)    | 0(%0)    | 0(%0) |            |
|                                                      | Arabic   | 50(%83.3) | 7(%11.6)  | 3(%5)    | 0(%0) |            |

pregnant women. In this study, ChatGPT-4 responses to 15 commonly asked pregnancy-related questions in three languages were evaluated for accuracy, empathetic tone, clarity and comprehensibility, and appropriate referral to obstetric care. Although performance was similar across languages in terms of referrals to obstetricians, significant differences were found in response accuracy, empathetic tone, and clarity.

Individual analysis showed that language-specific differences were present in several topics. For the question on pregnancy and nutrition, the English response did not mention herbal teas, carbonated drinks, or sweets, whereas the Turkish response included herbal teas but lacked details on dosage or consumption limits. The Turkish version was also more comprehensive, as it addressed probiotics and portion control. For signs of labor onset, the English response included additional details, such as hospital bag essentials, the “5-1-1 rule,” and Braxton Hicks contractions, which were absent or less detailed in the other languages. Similarly, when comparing spontaneous vaginal delivery and cesarean section, the English response mentioned pelvic organ prolapse and incontinence, while these important points were missing in the other languages. In the Arabic response on prenatal screening tests, specific tests such as NIPT, CVS, and amniocentesis were not mentioned. For supplements and vitamins during pregnancy, the English response provided a broader nutritional guide, whereas the Turkish and Arabic versions mainly focused on supplements.

Differences were also observed in lifestyle-related questions. Regarding exercise during pregnancy, the English response mentioned the risk of vena cava compression during supine exercises, a detail not included in the other languages. For relaxation methods such as massage and yoga, all versions were adequate; however, the English response added information on aromatherapy and warm baths, extending beyond the original question. For weight gain during pregnancy, the Arabic response was more detailed, whereas the Turkish version did not address special situations such as twin pregnancies. For fasting during pregnancy, the English response emphasized avoiding fasting, while the Arabic version discussed trimester-specific considerations and stated that fasting may be safe in the third trimester for healthy pregnant women. Although all versions warned that prolonged fasting or dehydration may harm both mother and baby, the Turkish response

additionally offered alternatives and mentioned religious flexibilities, whereas the Arabic version did not.

Overall, although ChatGPT responses appeared generally sufficient, detailed analysis revealed meaningful language-based differences in content, emphasis, and interpretation.

ChatGPT achieved over 80% success, defined as responses rated “excellent” or “good,” in the response accuracy category across all languages for frequently asked pregnancy-related questions. However, the model emphasized different aspects in each language. In previous studies, ChatGPT achieved an accuracy rate of 88% in providing breast cancer recommendations (18). It also outperformed newly graduated residents in gynecology and obstetrics in a virtual examination (19). In contrast, in an analysis of neurosurgery questions, AI achieved 54.9% accuracy across 477 non-image-based board-style questions (20). These findings indicate that ChatGPT may produce varying results across medical specialties (21,22). However, these studies were conducted in a single language and did not compare performance across languages.

When only responses rated as “excellent” were considered for medical accuracy, English responses had the highest proportion, reaching 95%. This finding may be related to the predominance of English-language content in model training datasets. Therefore, the performance of AI-based language models may vary depending on the scope and depth of training data available in each language.

Although significant differences were found in empathetic tone and clarity and comprehensibility across languages, overall performance was successful, with more than 95% of responses rated as “excellent” or “good.” However, ChatGPT’s responses exceeded the recommended 6<sup>th</sup>- to 8<sup>th</sup>-grade reading level for scientific texts, potentially limiting usability for individuals with lower literacy skills (23). In our study, the stronger performance of English responses may be related to the larger amount of English-language data used during model training. Conversely, the lower proportion of Arabic responses rated as “excellent” may reflect the more limited availability of Arabic data and the complexity of its linguistic structure. Nevertheless, the lack of a significant difference in appropriate referral to an obstetrician suggests that this function operates consistently across languages despite linguistic variation.

Overall, ChatGPT’s responses to pregnancy-related questions in three languages were generally rated

“good” or higher, suggesting it may serve as a useful source of health information for expectant mothers. It may be particularly helpful for women who are hesitant to ask certain questions, have limited access to healthcare professionals, or need immediate guidance when experiencing unusual symptoms. However, ChatGPT has limitations in interpreting some questions and may occasionally provide incorrect information (24). Therefore, it should be used cautiously and should not be considered a replacement for an obstetrician or other healthcare professional. Future studies should focus on improving ChatGPT’s medical knowledge and standardizing its performance across languages. Comparative studies in additional languages may provide more detailed insights, and once such standardization is achieved, ChatGPT could become a supportive tool for obstetricians and healthcare professionals by providing reliable information to expectant mothers.

The findings of this study represent an important step toward understanding performance variations of large language models across languages. The reliability and accuracy of AI-powered tools are particularly important in sensitive fields such as healthcare. Therefore, improving response quality across languages requires broader and more balanced training datasets, as well as critical evaluation of cross-language generalizability.

### Limitations

Although the evaluation criteria were assessed by expert obstetricians, some subjectivity may remain. In addition, the study included only the three most commonly used languages in our country; further research is needed to assess performance in additional languages. Another limitation is the restricted number of questions, which was limited to fifteen. Future studies with larger question sets may provide more comprehensive results.

### Conclusion

This study demonstrated that ChatGPT provides medically grounded responses to questions frequently asked by expectant mothers in three languages and may offer accessible information at any time to individuals with limited prior knowledge. Although the responses generally followed similar outlines, they varied across languages. Further research is needed to define the limitations of ChatGPT in the medical field across different languages (24). Multilingual studies, particularly in healthcare, will help evaluate and establish its effectiveness in each language.

### References

1. Zhang J, Lu JD, Chen B, Pan S, Jin L, Zheng Y, et al. Vision transformer introduces a new vitality to the classification of renal pathology. *BMC Nephrol.* 2024;25(1):337. <https://doi.org/10.1186/s12882-024-03800-x>
2. Chen Z, Yadollahpour A. A new era in cognitive neuroscience: the tidal wave of artificial intelligence (AI). *BMC Neurosci.* 2024;25(1):23. <https://doi.org/10.1186/s12868-024-00869-w>
3. Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakin S, Mim MMJ. A review on large language models: architectures, applications, taxonomies, open issues and challenges. *IEEE Access.* 2024;12:26839–26874. <https://doi.org/10.1109/ACCESS.2024.3365742>
4. Porto BM. Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review. *BMC Emerg Med.* 2024;24(1):219. <https://doi.org/10.1186/s12873-024-01135-2>
5. Chung P, Fong CT, Walters AM, Yetisgen M, O'Reilly-Shah VN. Prediction of American Society of Anesthesiologists physical status classification from preoperative clinical text narratives using natural language processing. *BMC Anesthesiol.* 2023;23(1):296. <https://doi.org/10.1186/s12871-023-02248-0>
6. Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, et al. A survey of large language models. *arXiv:2303.18223 [cs.CL].* <https://doi.org/10.48550/arXiv.2303.18223>
7. Gulati V, Roy SG, Moawad A, Garcia D, Babu A, Pott JD, et al. Transcending language barriers: can ChatGPT be the key to enhancing multilingual accessibility in healthcare? *J Am Coll Radiol.* 2024;21(12):1888–1895. <https://doi.org/10.1016/j.jacr.2024.05.009>
8. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. *Am J Obstet Gynecol.* 2023;228(6):696–705. <https://doi.org/10.1016/j.ajog.2023.03.009>
9. Chang Y-H, Lin Y-C, Huang F-W, Chen D-M, Chung Y-T, Chen W-K, et al. Using machine learning and natural language processing in triage for prediction of clinical disposition in the emergency department. *BMC Emerg Med.* 2024;24:237. <https://doi.org/10.1186/s12873-024-01152-1>
10. Chiarello F, Giordano V, Spada I, Barandoni S, Fantoni G. Future applications of generative large language models: a data-driven case study on ChatGPT. *Technovation.* 2024;133:103002. <https://doi.org/10.1016/j.technovation.2024.103002>
11. Briganti G. How ChatGPT works: a mini review. *Eur Arch Otorhinolaryngol.* 2024;281(3):1565–1569. <https://doi.org/10.1007/s00405-023-08337-7>
12. Allen MR, Webb S, Mandvi A, Freiden M, Ming T-S, Kallenberg G. Navigating the doctor-patient-AI relationship-a mixed-methods study of physician attitudes toward artificial intelligence in primary care. *BMC Prim Care.* 2024;25(1):42. <https://doi.org/10.1186/s12875-024-02282-y>

13. Wolfram S. WHAT is ChatGPT Doing ... and Why Does It Work? Wolfram Media, Inc.; 2023. 112 p. <https://doi.org/10.31855/bc47ee6b-75c>
14. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. *BMC Pregnancy Childbirth*. 2016;16:1–10. <https://doi.org/10.1186/s12884-016-0856-5>
15. Çelik E, Turgut MA, Aydoğan M, Kılınc M, Toktaş İ, Akelma H. Comparison of AI applications and anesthesiologist's anesthesia method choices. *BMC Anesthesiol*. 2025;25:2. <https://doi.org/10.1186/s12871-024-02882-2>
16. Han H, Li R, Fu D, Zhou H, Zhan Z, Wu Y et al. Revolutionizing spinal interventions: a systematic review of artificial intelligence technology applications in contemporary surgery. *BMC Surg*. 2024;24(1):345. <https://doi.org/10.1186/s12893-024-02646-2>
17. Serapio A, Chaudhari G, Savage C, Lee YJ, Vella M, Sridhar S, et al. An open-source fine-tuned large language model for radiological impression generation: a multi-reader performance study. *BMC Med Imaging*. 2024;24(1):254. <https://doi.org/10.1186/s12880-024-01435-w>
18. Haver HL, Ambinder EB, Bahl M, Oluyemi ET, Jeudy J, Yi PH. Appropriateness of breast cancer prevention and screening recommendations provided by ChatGPT. *Radiology*. 2023;307(4):e230424. <https://doi.org/10.1148/radiol.230424>
19. Li SW, Kemp MW, Logan SJ, Dimri PS, Singh N, Mattar CNZ, et al. ChatGPT outscored human candidates in a virtual objective structured clinical examination in obstetrics and gynecology. *Am J Obstet Gynecol*. 2023;229(2):172.e1–72. <https://doi.org/10.1016/j.ajog.2023.04.020>
20. Hopkins BS, Nguyen VN, Dallas J, Texakalidis P, Yang M, Renn A, et al. ChatGPT versus the neurosurgical written boards: a comparative analysis of artificial intelligence/machine learning performance on neurosurgical board-style questions. *J Neurosurg*. 2023;139(3):904–911. <https://doi.org/10.3171/2023.2.JNS23419>
21. Davies NP, Wilson R, Winder MS, Tunster SJ, McVicar K, Thakrar S, et al. ChatGPT sits the DFPH exam: large language model performance and potential to support public health learning. *BMC Med Educ*. 2024;24(1):57. <https://doi.org/10.1186/s12909-024-05042-9>
22. Zong H, Li J, Wu E, Wu R, Lu J, Shen B. Performance of ChatGPT on Chinese national medical licensing examinations: a five-year examination evaluation study for physicians, pharmacists and nurses. *BMC Med Educ*. 2024;24(1):143. <https://doi.org/10.1186/s12909-024-05125-7>
23. DeMarco J, Nystrom M. The importance of health literacy in patient education. *J Consum Health Internet*. 2010;14(3):294–301. <https://doi.org/10.1080/15398285.2010.502021>
24. Ye Z, Zhang B, Zhang K, Mendez MJG, Yan H, We T, et al. An assessment of ChatGPT's responses to frequently asked questions about cervical and breast cancer. *BMC Womens Health*. 2024;24(1):482. <https://doi.org/10.1186/s12905-024-03320-8>