

EVALUATION OF THE CHATGPT-4 ARTIFICIAL INTELLIGENCE RESPONSE TO THE PROBLEMS OF HIP REPLACEMENT

CHAT GPT-4 Yapay Zekanın Kalça Protezi Sorunlarına Cevabının Değerlendirilmesi

Hacı Ali OLÇAR¹, Gökçen AYDIN AKBUĞA², Samet ULUTÜRK³, Şenay ARAS DOĞAN², Murat KORKMAZ¹, İrem Nisa KORKMAZ⁴

ABSTRACT

Objective: This study aims to evaluate the CHAT GPT-4 artificial intelligence responses to common problems experienced by patients after hip replacement surgery using expert opinion.

Material and Methods: The study used an 11-question form developed by the researcher from a literature review of the information needs of patients after hip replacement surgery. Ten orthopaedic specialists rated the accuracy of the responses generated by Chat GPT-4 using a 5-point Likert scale (5: strongly agree, 1: strongly disagree). The maximum score that can be obtained from the rating scale is 55, and the minimum score is 11. The scores are summarised using descriptive statistics.

Results: All of the experts agreed that the artificial intelligence (AI) correctly answered the signs of wound infection (4.80±0.42), practices to prevent home accidents (4.50±0.52), and in which cases a doctor should be consulted (4.50±0.52). Half of the experts report that the artificial intelligence answered the question (Antithrombotic treatment time) partially true/false. For the question (Will my hip replacement dislocate?) 60 % of the experts report that the artificial intelligence answered partially true/false.

Conclusion: When Chat GPT-4 was evaluated by experts, it was found to provide mostly accurate information. However, it is thought that it may be more limited in individualised treatment and care practices. More research is needed to confirm the accuracy of the results.

Keywords: CHAT GPT-4; Hip Prosthesis; Patient Education

ÖZET

Amaç: Bu çalışma ile kalça protezi sonrası hastalarda sık görülen sorunlara yönelik CHAT GPT-4 yapay zeka cevaplarının uzman görüşüyle değerlendirilmesi amaçlanmaktadır.

Gereç ve Yöntemler: Çalışmada araştırmacı tarafından kalça protezi ameliyatı sonrası hastaların bilgi gereksinimlerine yönelik literatür taranarak oluşturulan 11 soruluk form kullanılmıştır. On ortopedi uzmanı Chat GPT-4 tarafından oluşturulan yanıtların doğruluğunu 5'li likert tipte bir puan skalası (5: Kesinlikle katılıyorum, 1: Kesinlikle Katılmıyorum) ile değerlendirmiştir. Puan skalasından alınabilecek maksimum puan 55, minimum puan 11 olarak belirtilmiştir. Puanlar, tanımlayıcı istatistiklerle özetlenmiştir.

Bulgular: Uzmanların tamamı, yapay zekanın yara yeri enfeksiyonu belirtilerini (4,80±0,42), ev kazalarını önlemeye yönelik uygulamaları (4,50±0,52), hangi durumlarda doktora başvurulacağını (4,50±0,52) doğru cevapladığını belirtmiştir. Uzmanların %50'si, yapay zekanın antitrombolitik tedavi süresini (3,80±1,39), %60'ı ise protezim çıkar mı? (3,20±0,91) sorusunu kısmen doğru/yanlış cevapladığını ifade etmiştir. Puan skalasından alınan toplam puan ortalaması 44,40±3,83 olarak bulunmuştur.

Sonuç: Chat GPT'nin uzmanlar tarafından değerlendirildiğinde büyük ölçüde doğru bilgiler verdiği belirlenmiştir. Ancak bireyselleştirilmiş tedavi ve bakım uygulamalarında daha kısıtlı kalabileceği düşünülmektedir. Sonuçların doğruluğunu kanıtlamak için daha fazla çalışmaya ihtiyaç duyulmaktadır.

Anahtar Kelimeler: Chat GPT-4; Kalça Protezi; Hasta Eğitimi

¹Yozgat Bozok Üniversitesi,
Ortopedi ve Travmatoloji Anabilim Dalı
²Yozgat Bozok Üniversitesi,
Sağlık Bilimleri Fakültesi,
Hemşirelik Bölümü.
³Yozgat Bozok Üniversitesi,
İletişim Fakültesi
⁴Bakırçay Üniversitesi,
Anatomi Bölümü

Hacı Ali OLÇAR, Dr. Öğr. Ü.
(0000-0002-3097-3677)
Gökçen AYDIN AKBUĞA, Doç. Dr.
(0000-0003-3839-7317)
Samet ULUTÜRK, Bilgisayar Mühendisi
(0000-0002-7765-6281)
Şenay ARAS DOĞAN, Dr. Öğr. Ü.
(0000-0002-0064-8929)
Murat KORKMAZ, Prof. Dr.
(0000-0002-5920-0280)
İrem Nisa KORKMAZ, Fizyoterapist
(0009-0005-6839-2869)

İletişim:
Dr. Öğr. Ü. Hacı Ali OLÇAR
Yozgat Bozok Üniversitesi, Ortopedi ve
Travmatoloji Anabilim Dalı

Geliş tarihi/Received: 22.06.2023
Kabul tarihi/Accepted: 26.10.2023
DOI: 10.16919/bozoktip.1318421

Bozok Tıp Derg 2023;13(4):118-122
Bozok Med J 2023;13(4):118-122

INDUCTION

Chat bots have attracted a lot of attention recently. The intended use of chat bots has been primitive and limited to answering rudimentary questions on help desk pages or addressing customer issues. However, Chat GPT-4, designed as a chat bot, is slowly ushering in a new era in the world of Natural Language Processing (NLP) with its ability to carry on a conversation over multiple queries and generate software code (1, 2). The difference is that while most chat bots consist of pre-programmed automatic responses, Chat GPT-4 is able to respond directly to the query and adapt to the chat like a human (3).

Chat GPT-4 is an artificial intelligence (AI) chatbot developed by Open AI and released in November 2022 (3). As a machine learning model, Chat GPT-4 is capable of supporting a wide range of NLP tasks. It is a next-generation artificial intelligence technology that can understand and generate human-like answers to a wide range of questions and requests, thanks to its training based on a large text dataset (4). Some of the potential benefits of Chat GPT-4 are:

- Ability to answer questions
- Solve mathematical equations,
- Ability to write texts (academic, literary articles, etc.)
- Ability to translate between languages,
- Ability to debug and correct,
- Ability to summarise text,
- Finding key words in texts,
- Ability to classify
- Ability to make suggestions
- Being able to explain what something does
- Developing software.

Google currently provides most of the functionality to the users. It also has the potential to perform tasks such as creating content for websites and answering customer questions. Like most artificial intelligence, Chat GPT-4 is trained using a large database of information from the web. It has a conversational interface that allows it to answer follow-up questions, reject certain requests, and recognise when it is malfunctioning or failing. Chat GPT-4 can compile and respond to the information it learns from the Internet, which is its database. The information it obtains is

actually accessible via the Google search engine. The difference is that while Google offers the user all the options related to the searched topic, i.e. different sources, Chat GPT-4 makes a series of guesses and presents a single answer to the user much faster.

Total Hip Replacement (THR) is one of the most difficult, complex and time-consuming procedures in orthopaedic surgery. Today, the number of THR operations is gradually increasing in parallel with the increase in life expectancy and the increase in musculoskeletal problems (osteoarthritis, rheumatoid arthritis, trauma, etc.) (5).

The aim of this study is to evaluate Chat GPT-4's artificial intelligence responses to common problems experienced by patients after hip replacement surgery with expert opinion.

MATERIALS AND METHODS

The questions in the study; reasons for repeated hospitalisation and admission, most common complications after THR, needs of patients at discharge. The questions were developed by the researcher through a literature review (6-11).

The study used an 11-question questionnaire developed by the researcher from a literature review of patients' information needs after hip replacement surgery (6-11). Ten orthopaedic specialists rated the accuracy of the responses generated by Chat GPT-4 using a 5-point Likert-type scale (5: strongly agree, 1: strongly disagree) (Table 1).

Our study is not a patient study, but a comparison of computer data with expert information. Therefore, ethics committee approval is not required.

The maximum score that can be obtained from the scale is 55 and the minimum score is 11. Scores are summarised with mean, standard deviation and percentage distribution.

The reliability of the data provided by Chat CPT-4 and its compatibility with the clinicians were assessed by scoring the responses of Chat CPT-4 to the questions posed by the orthopaedic and traumatology physicians.

RESULTS

All experts agreed that the AI correctly identified the signs of wound infection (4.80 ± 0.42), practices to prevent home accidents (4.50 ± 0.52), and in which

cases a doctor should be consulted (4.50±0.52). Half of the experts report that the artificial intelligence answered the question (Antithrombotic treatment time) partially true/false. For the question (Will my hip replacement dislocate?) 60 % of the experts report that the artificial intelligence answered partially true/false.

DISCUSSION

CHAT - GPT- 4 answered total 11 questions. Although the answer to the second question explains the dressing technique, it does not state that the person applying the dressing should be a healthcare professional.

Although the answer to the fifth question states that it is possible to start 4-6 weeks after the operation, it does not explain the need to assess rehabilitation in relation to the patient and the changes in the process of starting rehabilitation.

Although the answer to the sixth question states that it will continue for 10 to 14 days after the operation, in most clinics the treatment will continue for 1 month. Trained with information from the Internet, Chat GPT-4 builds sentences word by word, choosing the most likely phrase to follow each word. For this reason, Chat GPT-4 may give incorrect information from time to time. This is because the robot gathers information from the Internet without checking the reliability of the source.

Although various complications related to the prosthesis are explained in the answer to the eighth question, it is noticeable that the movements associated with prosthesis dislocation, which are shown in the aetiology of the prosthesis, and hip joint movements that should not be performed after the prosthesis, are not mentioned.

The shortcoming of the system was that it did not ask about gender and did not suggest gender-specific exercises in the answer to the eleventh question. In the current model, Chat GPT-4 guesses when the user asks a vague question, but does not ask clarifying questions. In our study, Chat CPT-4, which has such shortcomings and is one of the representatives of artificial intelligence technology, was evaluated in relation to prosthetic surgeries, which have an important place in human health. Our results showed that Chat CPT-4, despite its shortcomings, gave reasonable answers.

Chat GPT-4's capabilities are remarkable, but like many chat bots of the past, it has its own limitations and problems:

Artificial intelligence produces content with the data it receives from the Internet, but it is necessary to use

Table 1. Questions and Score scale

Please evaluate the CHAT GPT-4 results according to the table below					
	Absolutely agree	Agree	Partially Agree	disagree	strongly disagree
	5	4	3	2	1
1- What are the symptoms of wound infection?					
2- How should I dress the wound?					
3- What methods can I use to cope with pain?					
4- My pain increases with movement, what should I do?					
5- When can I exercise?					
6- How long will the antithrombolytic therapy continue?					
7- How should I organize my home to prevent accidents?					
8- Will my prosthesis come off?					
9- Does an animal-based diet accelerate my healing process?					
10- In which cases can I consult my doctor?					
11-When can I have sexual intercourse after hip replacement surgery?					

Table 2. Score Distribution of the Answers Given by CHAT GPT-4 According to Expert Opinions

Questions	minimum	Maximum	Mean ±SS
Question 1	4.00	5.00	4.80±0.42
Question 2	2.00	5.00	3.70±1.05
Question 3	3.00	5.00	4.10±0.56
Question 4	3.00	5.00	4.20±0.78
Question 5	2.00	4.00	3.50±0.70
Question 6	2.00	6.00	3.80±1.39
Question 7	4.00	5.00	4.50±0.52
Question 8	2.00	4.00	3.20±0.91
Question 9	3.00	5.00	4.20±0.63
Question 10	4.00	5.00	4.50±0.52
Question 11	3.00	5.00	3.90±0.56
Total	40.00	52.00	44.40±3.83

control tools to prevent plagiarism of the text. Chat GPT-4 sometimes responds with plausible but inaccurate information, and there may be difficulties in training Reinforcement Learning (RL) to correct information with this information. On the other hand, if it is trained to be more cautious, it may reject questions that it can answer correctly. That's because it's more about what the model knows than what the person knows. Chat GPT-4's response may be inconsistent if a small change is made to the data entry or if the same question is asked several times. Due to bias and over-optimization in the training data, the Chat GPT-4 model may be overly verbose and overuse certain phrases (12-13).

CONCLUSION

The evaluation of artificial intelligence, such as ChatGPT-4, in addressing issues related to hip replacement surgery is a significant and evolving area of research and application in the field of healthcare. AI systems like ChatGPT-4 can play a crucial role in providing preoperative and postoperative information and education to patients undergoing hip replacement surgery. They can explain the procedure, potential risks, expected outcomes, and post-operative care instructions, ensuring patients are well informed and mentally prepared.

Our study upon comprehensive evaluation by experts, it was discerned that Chat GPT typically furnishes accurate information, yet apprehensions have arisen

regarding its potential limitations in catering to individualized treatment and care scenarios.

It is important to view AI as a helpful tool for healthcare professionals rather than a replacement for them. The evaluation of AI, like ChatGPT-4, in addressing hip replacement surgery issues demonstrates the potential for significant improvements in patient care, surgical precision, and postoperative monitoring. However, we also underscores the need for careful implementation, ethical considerations, and ongoing research to harness AI is full potential in the field of orthopedic surgery.

There is a growing consensus within the scientific community that additional research is imperative to ensure precise validation the precision of these initial findings.

Acknowledgement

There is no conflict of interest between the authors. Also there is not any funding for the study .

KAYNAKLAR

1. Teubner, T., Flath, C. M., Weinhardt, C., van der Aalst, W., & Hinz, O. (2023). Welcome to the era of chatgpt et al. the prospects of large language models. *Business & Information Systems Engineering*, 65(2), 95-101.
2. Birer, G. C., "ChatGPT", TÜBİTAK Bilim ve Teknik Dergisi, 2023;662:36-7
3. Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2023). Chatting and

cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 1-12.

4. Daniel, G., Cabot, J., Deruelle, L., & Derras, M. Xatkit: a multimodal low-code chatbot development framework. *IEEE Access*, 8, 15332-46.

5. OECD. Hip and knee replacement. In *Health Care Activities Health at a Glance 2015: OECD Indicators*. OECD Publishing, Paris, 2015; 112-3.

6. Balkan A, Seki Z. Determination of Complications of Patients Undergoing Total Hip Replacement after Discharge. *BSJ Health Sci*, 2021; 4(2): 141-9.

7. Lucas B. Total hip and total knee replacement: postoperative nursing management. *Br J Nurs*. 2008 Dec 11-2009 Jan 7;17(22):1410-4

8. Yıldırım A, Şendir M. Postoperative Pain Assessment and Pain Management in Patients Undergoing Total Knee Replacement. *University of Health Sciences Journal of Nursing*. 2013; 1(3): 157-64.

9. Montin, L., Suominen, T., & Leino-Kilpi, H. The experiences of patients undergoing total hip replacement. *Journal of Orthopaedic Nursing*, 2002;6(1):23-9.

10. Kaya, Ç., & Bilik, Ö. Effect of Counseling on Quality of Life and Self-Care Agency for Patients Who are Scheduled for Total Knee Replacement. *Clinical Nursing Research*, 2022;31(3):519-29.

11. Kuru, T., & Olcar, H. A. Effects of early mobilization and weight bearing on postoperative walking ability and pain in geriatric patients operated due to hip fracture: a retrospective analysis. *Turkish journal of medical sciences*, 2020;50(1):117-25.

12. Howard A, Hope W, Gerada A. ChatGPT and antimicrobial advice: the end of the consulting infection doctor? *Lancet Infect Dis*. 2023 Apr;23(4):405-6.

13. La Grandeur, K. (2021). How safe is our reliance on AI, and should we regulate it? *AI and Ethics*, 1, 93-9.