



ChatGPT's potential to support home care for patients in the early period after orthopedic interventions and enhance public health

Dilek Yapar, MD^{1,2}, Yasemin Demir Avcı, PhD^{2,3}, Esra Tokur Sonuvar, MSc²,
Ömer Faruk Eğerci, MD⁴, Aliekber Yapar, MD⁴

¹Department of Public Health, Turkish Ministry of Health, Muratpaşa District Health Directorate, Antalya, Türkiye

²Institute of Health Science, Medical Informatics, Akdeniz University, Antalya, Türkiye

³Department of Public Health Nursing, Akdeniz University Faculty of Nursing, Türkiye

⁴Department of Orthopedics and Traumatology, Antalya Training and Research Hospital, Antalya, Türkiye

Artificial intelligence (AI), with its diverse applications across sectors, including healthcare, education, and finance, has brought groundbreaking changes to numerous fields.^[1,2] A notable offshoot, natural language processing, empowers computers to understand and produce human language. Among natural language processing tools, large language models stand out. These models, particularly OpenAI's (OpenAI, Inc., San Francisco, CA, USA) GPT (Generative Pre-Trained Transformer) series, culminating in GPT-4 in 2023, utilize deep learning to generate human-like text, revolutionizing interfaces such as chatbots.^[3-6] Its capabilities span from

ABSTRACT

Objectives: This study presents the first investigation into the potential of ChatGPT to provide medical consultation for patients undergoing orthopedic interventions, with the primary objective of evaluating ChatGPT's effectiveness in supporting patient self-management during the essential early recovery phase at home.

Materials and methods: Seven scenarios, representative of common situations in orthopedics and traumatology, were presented to ChatGPT version 4.0 to obtain advice. These scenarios and ChatGPT's responses were then evaluated by 68 expert orthopedists (67 males, 1 female; mean age: 37.9±5.9 years; range, 30 to 59 years), 40 of whom had at least four years of orthopedic experience, while 28 were associate or full professors. Expert orthopedists used a rubric on a scale of 1 to 5 to evaluate ChatGPT's advice based on accuracy, applicability, comprehensiveness, and clarity. Those who gave ChatGPT a score of 4 or higher considered its performance as above average or excellent.

Results: In all scenarios, the median evaluation scores were at least 4 across accuracy, applicability, comprehensiveness, and communication. As for mean scores, accuracy was the highest-rated dimension at 4.2±0.8, while mean comprehensiveness was slightly lower at 3.9±0.8. Orthopedist characteristics, such as academic title and prior use of ChatGPT, did not influence their evaluation (all p>0.05). Across all scenarios, ChatGPT demonstrated an accuracy of 79.8%, with applicability at 75.2%, comprehensiveness at 70.6%, and a 75.6% rating for communication clarity.

Conclusion: This study emphasizes ChatGPT's strengths in accuracy and applicability for home care after orthopedic intervention but underscores a need for improved comprehensiveness. This focused evaluation not only sheds light on ChatGPT's potential in specialized medical advice but also suggests its potential to play a broader role in the advancement of public health.

Keywords: ChatGPT, expert evaluation, large language models, medical consultation, orthopedic interventions, rubric.

Received: August 27, 2023

Accepted: November 06, 2023

Published online: November 30, 2023

Correspondence: Aliekber Yapar, MD. Antalya Eğitim ve Araştırma Hastanesi, Ortopedi ve Travmatoloji Kliniği, 07100 Muratpaşa, Antalya, Türkiye.

E-mail: aliekberyapar@hotmail.com

Doi: 10.52312/jdrs.2023.1402

Citation: Yapar D, Demir Avcı Y, Tokur Sonuvar E, Eğerci ÖF, Yapar A. ChatGPT's potential to support home care for patients in the early period after orthopedic interventions and enhance public health. Jt Dis Relat Surg 2024;35(1):169-176. doi: 10.52312/jdrs.2023.1402.

©2024 All right reserved by the Turkish Joint Diseases Foundation

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes (<http://creativecommons.org/licenses/by-nc/4.0/>).

analyzing patient data to understanding complex medical literature, offering health information, and improving text writing, indicating the promising potential of future GPT versions.^[7-14] Furthermore, ChatGPT can improve health service accessibility and quality, particularly for patients in remote areas, by providing medical information and aiding in the comprehension of complex medical data, thus facilitating informed decisions.^[4,13,15] Thus, investigating ChatGPT's capacity to offer medical consultation represents a significant stride in potentially elevating public health's overall quality and accessibility.^[16,17] Such technological innovations could enhance key aspects of public health, including accessibility, information dissemination, patient awareness, and cost-effectiveness of health services. Recently, some journals have accepted and published case reports created with the assistance of ChatGPT, explicitly acknowledging ChatGPT's contribution in the titles and acknowledgment sections.^[5] These examples demonstrate how AI can play an effective role in generating and publishing scientific articles. Another example that concretely illustrates the potential contribution of AI to medical research is the appearance of ChatGPT as a coauthor in publications. Despite the risk of bias and inaccuracies in AI-generated articles, the question of whether AI can be considered an author has sparked debates.^[7,18,19]

Orthopedic interventions, among the most frequent surgical procedures, are vital for restoring mobility and enhancing the quality of life for patients. The home-care phase is essential for recovery but can be challenging due to complex medical guidelines and differing levels of patient understanding.^[20,21] Effective self-management, guided by clear instructions, is key to swift recovery and minimizing complications in the home-care phase. Given this context, there arises a need for innovative solutions that can bridge the information gap. ChatGPT emerges as a promising AI tool with the potential to provide patients with timely information and support.^[19] Recently, studies emphasizing ChatGPT's capability in orthopedic knowledge acquisition have started to emerge. Kaarre et al.^[22] demonstrated ChatGPT's capability to respond to inquiries related to anterior cruciate ligament surgery and contribute to acquiring orthopedic knowledge. Another study emphasized ChatGPT's potential to improve patient education and engagement by serving as a virtual assistant, offering patients relevant information regarding their orthopedic conditions, treatment choices, and postoperative care.^[23] There is a need for expert evaluations to understand ChatGPT's potential

and feasibility in providing medical consultation to orthopedic patients. The primary aim of this research was to explore ChatGPT's effectiveness in fostering patient self-management during the crucial early recovery phase at home.

MATERIALS AND METHODS

Case scenario development and ChatGPT's responses:

In this descriptive cross-sectional study, based on the most commonly encountered cases in orthopedics and traumatology, two orthopedists prepared seven scenarios covering the areas of arthroplasty, trauma, tumor, spine, hand surgery, pediatrics, sports injuries, arthroscopy, and the foot, ankle, shoulder, and elbow. In these scenarios, the researchers focused on cases of patients who either presented with orthopedic emergencies and received interventions or were discharged home after elective orthopedic surgery. Scenarios were presented to GPT-4 on June 20, 2023, and medical advice was obtained. To assess the accuracy and alignment of ChatGPT's recommendations with real-world orthopedic solutions, these scenarios, along with ChatGPT's responses, were sent to expert orthopedists via Google Forms (Google LLC, Mountain View, CA, USA).

Evaluation of ChatGPT responses

Orthopedic specialists were asked to assess ChatGPT's recommendations for various case scenarios based on a rubric. Rubrics are standardized evaluation tools used to gauge the quality of a specific performance or outcome. Although rubric evaluations are commonly used in educational fields, particularly for grading written assignments or projects, they are also applicable in research and various other domains, including health.^[24-26] The rubric encompassed criteria of accuracy, applicability, comprehensiveness, and clarity of language.

Accuracy demonstrates the medical accuracy of the recommendations rated on a scale of 1 to 5, with 1 being entirely inaccurate and 5 being entirely accurate. Applicability shows the feasibility and patient-friendliness of the advice rated from 1 to 5, with 1 being inapplicable and 5 being very applicable. Comprehensiveness demonstrates to what extent the response covers various aspects of patient care. It is rated on a scale of 1 to 5, with 1 covering a single aspect and 5 covering multiple aspects. Communication displays whether the advice is conveyed in a manner easily understood by patients. It is rated on a scale of 1 to 5, with 1 being unclear and 5 being very clear.

Higher scores from ChatGPT responses indicate reliable and effective recommendations for early home care following orthopedic interventions. The percentage of orthopedists with a Likert response ≥ 4 signifies above-average or excellent performance for each respective measurement in the rubric evaluation. This metric was chosen to provide a clearer perspective on the proportion of evaluators who deemed ChatGPT's responses as highly satisfactory across various case scenarios. The overall evaluation percentage represents the average percentage of orthopedic specialists who provided a rating ≥ 4 on the Likert scale across all scenarios and dimensions. These percentages were aggregated for each dimension and then divided by seven (the number of scenarios) to calculate the overall percentage. This metric offers a comprehensive view of ChatGPT's overall performance, underscoring its capability to generate reliable and effective recommendations in orthopedic cases.

Expert panel and characteristics of orthopedists performing the evaluation

Evaluators of this study were orthopedics and traumatology specialists aged ≥ 30 years with a minimum of four years of experience. Those over the age of 65 were excluded. The primary role of the evaluators was to assess the recommendations provided by ChatGPT for various case scenarios in terms of accuracy, applicability, comprehensiveness, and communication skills using a rubric method.

The number of experts needed for such evaluations depends on the originality and scope of the study. For content validity, a minimum of five experts is typically recommended; however, for more complex studies, this number can increase up to 40.^[27,28] Nonetheless, during the study, we managed to reach a total of 68 evaluators (67 males, 1 female; mean age: 37.9 ± 5.9 years; range, 30 to 59 years). Expert orthopedists working in different institutions and meeting the inclusion criteria were reached via email using a snowball sampling method.

Statistical analysis

The data was analyzed using IBM SPSS version 22.0 software (IBM Corp., Armonk, NY, USA). The normality of continuous variables was assessed using both visual (histogram and probability plots) and analytical methods (Kolmogorov-Smirnov/Shapiro-Wilk tests). Descriptive statistics were presented as frequency, percentages, mean \pm standard deviation, and median (min-max). Since the rubric evaluation scores did not follow a normal distribution, the Mann-Whitney U test was used to make comparisons between two independent groups. A p-value < 0.05 was considered statistically significant.

RESULTS

Fifty-nine percent ($n=40$) of expert orthopedists held a specialist title, while 41% ($n=28$) were associate professors or professors. In terms of familiarity with

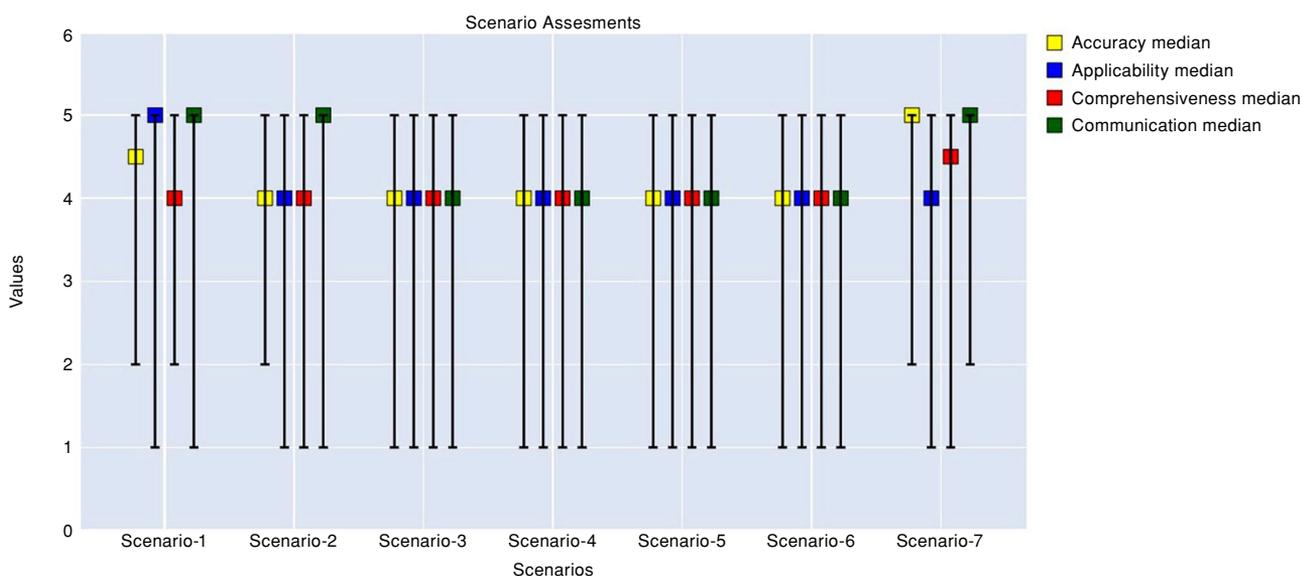


FIGURE 1. Range of Scores for dimensions of the scenarios.

TABLE I
Evaluating ChatGPT's responses to scenarios based on scoring rubric for accuracy, applicability, comprehensiveness, and communication (n=68)

Orthopedists	Orthopedists with Likert response of ≥ 4		Score		
	n	%	Mean \pm SD	Median	Min-Max
Scenario 1: Complications following cast treatment for forearm fracture in a 10-year-old child					
Accuracy	58	85.3	4.3 \pm 0.9	4.5	2-5
Applicability	54	79.4	4.3 \pm 1	5	2-5
Comprehensiveness	49	72.1	4 \pm 0.9	4	2-5
Communication	53	77.9	4.2 \pm 1	5	1.5
Scenario 2: Postoperative care concerns following knee replacement surgery in a 65-year-old patient					
Accuracy	50	73.5	4.1 \pm 1	4	2-5
Applicability	48	70.6	4 \pm 1	4	1-5
Comprehensiveness	46	67.6	3.9 \pm 1	4	1-5
Communication	52	76.5	4.2 \pm 1	5	1-5
Scenario 3: Postoperative care and guidance for a 17-year-old scoliosis patient					
Accuracy	57	83.8	4.2 \pm 1	4	1-5
Applicability	54	79.4	4.1 \pm 1	4	1-5
Comprehensiveness	51	75	4 \pm 1	4	1-5
Communication	52	76.5	4.1 \pm 1.1	4	1-5
Scenario 4: Postoperative care after hip replacement: Anticoagulants and compression stockings					
Accuracy	53	77.9	4.2 \pm 1	4	1-5
Applicability	47	69.1	3.9 \pm 1.1	4	1-5
Comprehensiveness	41	60.3	3.7 \pm 1.1	4	1-5
Communication	50	73.5	4 \pm 1.1	4	1-5
Scenario 5: Post-hip replacement precautions and care for a 60-year-old obese female patient					
Accuracy	54	79.4	4.1 \pm 0.9	4	1-5
Applicability	54	79.4	4.1 \pm 1	4	1-5
Comprehensiveness	50	73.5	3.9 \pm 1.1	4	1-5
Communication	49	72.1	4 \pm 1.1	4	1-5
Scenario 6: Post-surgery complications: Dealing with re-injury in arm fractures					
Accuracy	53	77.9	4 \pm 1	4	1-5
Applicability	50	73.5	4 \pm 1.1	4	1-5
Comprehensiveness	48	70.6	3.9 \pm 1.1	4	1-5
Communication	51	75	4.1 \pm 1	4	1-5
Scenario 7: Managing hip dysplasia in a 2-month-old infant: Questions and concerns					
Accuracy	55	80.9	4.3 \pm 0.9	5	2-5
Applicability	51	75	4.1 \pm 1	4	1-5
Comprehensiveness	51	75	4.1 \pm 1.1	4.5	1-5
Communication	53	77.9	4.2 \pm 0.9	5	2-5
All scenarios	Overall evaluation percentage*		Overall evaluation percentage**	Overall evaluation percentage**	
Accuracy	79.8		4.2 \pm 0.8	4.4	1.6-5
Applicability	75.2		4.1 \pm 0.9	4.4	1.4-5
Comprehensiveness	70.6		3.9 \pm 0.8	4.1	1.4-5
Communication	75.6		4.1 \pm 0.9	4.4	1.4-5

* The percentage of responses scoring ≥ 4 across all seven scenarios is provided for each of the four evaluation areas.

** Overall evaluation percentage for all scenarios is derived by aggregating and averaging the scores from all four areas across each scenario.

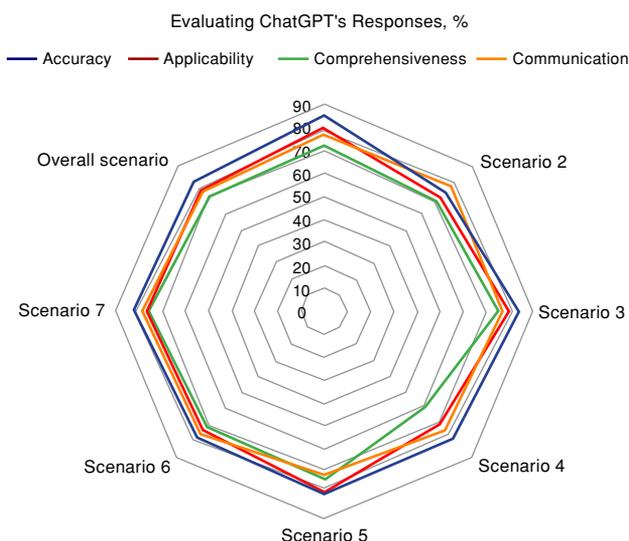


FIGURE 2. Distribution of scores ≥ 4 for scenarios across the four evaluation areas.

ChatGPT, 35.3% (n=24) of the orthopedists had no knowledge of it, 35.3% (n=24) had minimal knowledge, 22.1% (n=15) had basic knowledge, and the remaining 7.4% (n=5) had adequate or superior knowledge. Interestingly, only 45.6% of the orthopedists had previously used ChatGPT (n=31).

This study assessed the responses of ChatGPT to seven different medical scenarios using a Likert scoring rubric for accuracy, applicability, comprehensiveness, and communication. In all scenarios, the median evaluation scores were at least 4 (Figure 1). Excluding the comprehensiveness for the second scenario and both comprehensiveness and applicability for the fourth scenario, more than 70% of the orthopedists rated ChatGPT's responses ≥ 4 in all four areas across the other scenarios (Table I, Figure 2). Additionally, the overall index percentage of responses scoring ≥ 4 was calculated for each of the four evaluation dimensions (accuracy, applicability, comprehensiveness, and communication). For each dimension, the percentage of responses scoring ≥ 4 in each scenario was aggregated and then divided by seven. The resulting values ranged between 70 and 80% for the different dimensions (Table I). The highest mean ratings across all scenarios were observed for accuracy (4.2 ± 0.8), with comprehensiveness having the lowest mean score (3.9 ± 0.8 ; Table I).

In Table II, the overall index scores for ChatGPT's responses across all scenarios were compared based on different characteristics of

TABLE II Comparison of orthopedists' scores for ChatGPT responses based on evaluator characteristics					
	All Scenarios Overall Index Score*				p^{**}
	Median	Min-Max	Median	Min-Max	
Groups	Specialist (n=40)		Associate Professor/Professor (n=28)		
Accuracy	4.4	1.6-5	4.4	2.7-5	0.543
Applicability	4.2	1.4-5	4.6	2.7-5	0.102
Comprehensiveness	3.9	1.4-5	4.3	2.7-5	0.457
Communication	4.1	1.4-5	4.7	2.7-5	0.109
Groups	Using ChatGPT (n=31)		Not using ChatGPT (n=37)		
Accuracy	4.3	2-5	4.6	1.6-5	0.181
Applicability	4.3	1.9-5	4.4	1.4-5	0.198
Comprehensiveness	3.9	1.4-5	4.3	1.4-5	0.120
Communication	4.4	1.7-5	4.4	1.4-5	0.867
Groups	Age ≤ 35 years old (n=29)		Age > 35 years old (n=39)		
Accuracy	4.1	1.6-5	4.6	1.6-5	0.082
Applicability	4.1	1.9-5	4.6	1.4-5	0.124
Comprehensiveness	3.9	1.4-5	4.3	1.4-5	0.204
Communication	4.1	1.7-5	4.4	1.4-5	0.438

* The overall index scores for all scenarios is derived by aggregating and averaging the scores from all four areas across each scenario; ** Mann-Whitney-U test

the evaluating orthopedists. The characteristics considered were the orthopedist's academic title (specialist *vs.* associate professor/professor), whether they had used ChatGPT before, and their age (≤ 35 *vs.* > 35 years). Comparisons based on orthopedists' characteristics, such as academic title, usage of ChatGPT, and age, revealed no significant differences in evaluation scores across accuracy, applicability, comprehensiveness, or communication ($p > 0.05$ for all). This suggests that these factors do not influence the assessment of ChatGPT's performance in medical scenarios.

DISCUSSION

This study explores ChatGPT's potential in supporting patients during home care following orthopedic interventions. It presents expert evaluations of the AI-generated responses, specifically assessing their accuracy, applicability, comprehensiveness, and communication. This assessment was conducted by evaluating ChatGPT's responses to seven distinct medical scenarios using a Likert-type rubric. In our study, we examined a range of scenarios, from pediatric fractures to scoliosis surgeries, highlighting the comprehensive nature of our investigation. It was revealed that ChatGPT provides medical advice for potential home-care situations related to these scenarios with high accuracy scores. However, while the accuracy, applicability, and communication dimensions were highly rated, the comprehensiveness dimension received slightly lower scores. This could be attributed to the vast nature of medical information and the potential for nuanced details to be omitted in AI-generated responses. Large language models such as ChatGPT are powerful in processing and generating language, but they might not always provide comprehensive answers in complex medical situations.^[16]

The consistent evaluation scores across orthopedists, regardless of their academic standing, prior experience with ChatGPT, or age, indicate that ChatGPT provides universally understandable and standardized responses. In all presented scenarios, median evaluation scores were > 4 , showing orthopedists' favorable view of ChatGPT's answers. In our evaluation covering multiple scenarios, ChatGPT's responses had an accuracy rate of 79.8% and exhibited 70.6% comprehensiveness. One study found that ChatGPT provided suitable answers for 84% of cardiovascular disease prevention queries.^[29] Another emphasized its empathetic approach and response quality.^[4] Another study on vaccination showed 85.4% clarity and accuracy.^[30] While AI

tools can provide useful healthcare insights when prompted correctly, there exists the risk of receiving misleading answers without expert supervision. It is essential to highlight that our scenarios were crafted by seasoned physicians, underscoring the notion that the quality of a query directly influences the accuracy of the response.

In today's digital age, with the surge of patient messages, there is a growing demand for digital accessibility from medical professionals.^[15] Artificial intelligence chatbots, such as ChatGPT, could offer potential solutions, providing timely responses and enhancing the healthcare experience. While this study sheds light on the potential of ChatGPT in orthopedic home care, it is paramount to approach the findings with caution; AI tools should complement, not replace, human expertise.^[31] Furthermore, it is imperative that these models are updated and trained with the latest medical data to ensure their advice remains relevant.^[4,17] ChatGPT's integration into the healthcare field brings about ethical and safety concerns. Numerous studies highlighted the potential risks, such as bias or the dissemination of misleading information.^[9,16,32-34] Addressing these risks is paramount, and our study offers insights that could assist in shaping policies and guidelines for these technologies' ethical use.

As AI continues to play a role in healthcare services, addressing ethical and security risks is of paramount importance. Considering the potential ethical consequences, limitations, and challenges of being entirely reliant on AI systems for home care after medical intervention, these applications should be used as complementary tools, tailored to specific patient groups based on need, rather than replacing physicians entirely.^[31] The predominant use of AI systems in home care could lead to patients self-managing their health without medical supervision, potentially resulting in misdiagnosis or incorrect treatment and raising ethical concerns.^[31] Additionally, there may be ambiguity regarding responsibility for erroneous or misleading results from these systems. Additionally, evaluations without medical supervision may lead to a lack of fulfillment of patients' emotional and psychological needs.^[31] Another concern is that individuals or regions with limited access to technology may not fully benefit from these systems' advantages, leading to ethical disparities.^[31]

This research pioneers the exploration of ChatGPT's capabilities in orthopedics. It not only underscores ChatGPT's versatility but also paves the way for future studies in related fields. Furthermore,

this evaluation suggests that ChatGPT has the potential to play a broader role in advancing public health beyond merely specialized medical advice.

In this study, the primary focus was on expert evaluation, and the results of the study have highlighted the necessity of incorporating patient assessments. This limitation has, in turn, laid the groundwork for a follow-up study aimed at providing a more comprehensive evaluation of ChatGPT's capabilities and real-world applicability by including patient perspectives. Additionally, in terms of addressing ethical concerns and potential challenges associated with AI systems in healthcare, this future study has gained even more significance.

In conclusion, ChatGPT has showcased significant potential in aiding home care for orthopedic patients by providing accurate and actionable medical recommendations. However, this study underscores the need for enhanced comprehensiveness. If tools such as ChatGPT are developed to fit the healthcare sector, patients can access accurate and reliable health information more swiftly and make more informed health decisions. Furthermore, such tools can play a pivotal role in enhancing public health by providing consistent and trustworthy guidance, particularly during times when access to health services is limited. However, the success of these tools is closely tied to patients' ability to communicate effectively with them and correctly use the platform. Future studies should consider integrating ChatGPT with other health platforms and testing its efficacy across varied patient groups. It is also essential to enhance user experience for wider adoption.

Ethics Committee Approval: The study protocol was approved by the Antalya Training and Research Hospital Clinical Research Ethics Committee (date: 08.06.2023, no: 8/11-2023). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Data Sharing Statement: The detailed scenarios and ChatGPT responses can be shared with readers upon their request.

Author Contributions: Idea/concept, control/supervision: D.Y., Y.D.A., E.T.S., A.Y.; Design, data collection and/or processing, literature review, references and fundings, materials: D.Y., Y.D.A., E.T.S., Ö.F.E., A.Y.; Analysis and/or interpretation, writing the article: D.Y., Y.D.A., E.T.S.; Critical review: Ö.F.E., A.Y.

Conflict of Interest: The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding: The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Amara A, Hadj Taieb MA, Ben Aouicha M. Network representation learning systematic review: Ancestors and current development state. *Mach Learn Appl* 2021;6:100130. doi: 10.1016/j.mlwa.2021.100130.
2. Atik OŞ. Artificial intelligence, machine learning, and deep learning in orthopedic surgery. *Jt Dis Relat Surg* 2022;33:484-5. doi: 10.52312/jdrs.2022.57906.
3. Adamopoulou E, Moussiades L. Chatbots: History, technology, and applications. *Mach Learn Appl* 2020;2:100006. doi: 10.1016/j.mlwa.2020.100006.
4. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med* 2023;183:589-96. doi: 10.1001/jamainternmed.2023.1838.
5. Zhou Z. Evaluation of ChatGPT's capabilities in medical report generation. *Cureus* 2023;15:e37589. doi: 10.7759/cureus.37589.
6. OpenAI. OpenI ChatGpt Guide. Chatgpt History: Timeline, Facts, Version, Current Capability 2023 Available at: <https://openichatgptguide.com/chatgpt-history-timeline-facts-versions/> [Accessed: 15.06.2023].
7. Ariyaratne S, Botchu R, Iyengar KP. ChatGPT in academic publishing: An ally or an adversary? *Scott Med J* 2023;68:129-30. doi: 10.1177/00369330231174231.
8. Arslan S. Exploring the potential of Chat GPT in personalized obesity treatment. *Ann Biomed Eng* 2023;51:1887-8. doi: 10.1007/s10439-023-03227-9.
9. Au Yeung J, Kraljevic Z, Luintel A, Balston A, Idowu E, Dobson RJ, et al. AI chatbots not yet ready for clinical use. *Front Digit Health* 2023;5:1161098. doi: 10.3389/fdgth.2023.1161098.
10. Baumgartner C. The potential impact of ChatGPT in clinical and translational medicine. *Clin Transl Med* 2023;13:e1206. doi: 10.1002/ctm2.1206.
11. Berger U, Schneider N. How ChatGPT will Change Research, Education and Healthcare? *PPmP* 2023;73:159-61.
12. OpenAI. ChatGPT. Available at: <https://openai.com/blog/chatgpt/> [Accessed: 15.06.2023].
13. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: An analysis of multiple clinical and research scenarios. *J Med Syst* 2023;47:33. doi: 10.1007/s10916-023-01925-4.
14. Fröhling L, Zubiaga A. Feature-based detection of automated language models: Tackling GPT-2, GPT-3 and Grover. *PeerJ Comput Sci* 2021;7:e443. doi: 10.7717/peerj-cs.443.
15. Nov O, Singh N, Mann D. Putting ChatGPT's medical advice to the (Turing) test: Survey study. *JMIR Med Educ* 2023;9:e46939. doi: 10.2196/46939.
16. Sallam M. ChatGPT utility in healthcare education, research, and practice: Systematic review on the promising perspectives and valid concerns. *Healthcare (Basel)* 2023;11:887. doi: 10.3390/healthcare11060887.
17. Ayers JW, Zhu Z, Poliak A, Leas EC, Dredze M, Hogarth M, et al. Evaluating artificial intelligence responses to public health questions. *JAMA Netw Open* 2023;6:e2317517. doi: 10.1001/jamanetworkopen.2023.17517.
18. Dahmen J, Kayaalp ME, Ollivier M, Pareek A, Hirschmann MT, Karlsson J, et al. Artificial intelligence bot ChatGPT

- in medical research: The potential game changer as a double-edged sword. *Knee Surg Sports Traumatol Arthrosc* 2023;31:1187-9. doi: 10.1007/s00167-023-07355-6.
19. Ashraf H, Ashfaq H. The role of ChatGPT in medical research: Progress and limitations. *Ann Biomed Eng* 2023. doi: 10.1007/s10439-023-03311-0.
 20. Punnoose A, Claydon-Mueller LS, Weiss O, Zhang J, Rushton A, Khanduja V. Prehabilitation for patients undergoing orthopedic surgery: A systematic review and meta-analysis. *JAMA Netw Open* 2023;6:e238050. doi: 10.1001/jamanetworkopen.2023.8050.
 21. Stoicea N, Magal S, Kim JK, Bai M, Rogers B, Bergese SD. Post-acute transitional journey: Caring for orthopedic surgery patients in the United States. *Front Med (Lausanne)* 2018;5:342. doi: 10.3389/fmed.2018.00342.
 22. Kaarre J, Feldt R, Keeling LE, Dadoo S, Zsidai B, Hughes JD, et al. Exploring the potential of ChatGPT as a supplementary tool for providing orthopaedic information. *Knee Surg Sports Traumatol Arthrosc* 2023;31:5190-8. doi: 10.1007/s00167-023-07529-2.
 23. Hernigou P, Scarlat MM. Two minutes of orthopaedics with ChatGPT: It is just the beginning; it's going to be hot, hot, hot! *Int Orthop* 2023;47:1887-93. doi: 10.1007/s00264-023-05887-7.
 24. Chen X, Acosta S, Barry AE. Machine or human? Evaluating the quality of a language translation mobile app for diabetes education material. *JMIR Diabetes* 2017;2:e13. doi: 10.2196/diabetes.7446.
 25. Khanna RR, Karliner LS, Eck M, Vittinghoff E, Koenig CJ, Fang MC. Performance of an online translation tool when applied to patient educational material. *J Hosp Med* 2011;6:519-25. doi: 10.1002/jhm.898.
 26. Moskal BM, Leydens JA. Scoring rubric development: Validity and reliability. *Pract Assess Res Evaluation* 2000;7. doi: 10.7275/q7rm-gg74.
 27. Ayre C, Scally AJ. Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Meas Eval Couns Dev* 2014;47:79-86. doi: 10.1177/0748175613513808.
 28. Yurdugül H. Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. XIV. Ulusal Eğitim Bilimleri Kongresi. 28-30 Eylül 2005, Denizli; 2005;1:771-4.
 29. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA* 2023;329:842-4. doi: 10.1001/jama.2023.1044.
 30. Deiana G, Dettori M, Arghittu A, Azara A, Gabutti G, Castiglia P. Artificial intelligence and public health: Evaluating ChatGPT responses to vaccination myths and misconceptions. *Vaccines (Basel)* 2023;11:1217. doi: 10.3390/vaccines11071217.
 31. Atik OŞ. Writing for Joint Diseases and Related Surgery (JDRS): There is something new and interesting in this article! *Jt Dis Relat Surg* 2023;34:533. doi: 10.52312/jdrs.2023.57916.
 32. Samaan JS, Yeo YH, Rajeev N, Hawley L, Abel S, Ng WH, et al. Assessing the accuracy of responses by the language model ChatGPT to questions regarding bariatric surgery. *Obes Surg* 2023;33:1790-6. doi: 10.1007/s11695-023-06603-5.
 33. Sharma P. Chatbots in medical research: Advantages and limitations of artificial intelligence-enabled writing with a focus on ChatGPT as an author. *Clin Nucl Med* 2023;48:838-9. doi: 10.1097/RLU.0000000000004665.
 34. Singh OP. Artificial intelligence in the era of ChatGPT - Opportunities and challenges in mental health care. *Indian J Psychiatry* 2023;65:297-8. doi: 10.4103/indianjpsychiatry.indianjpsychiatry_112_23.