Joint Diseases and Related Surgery

Jt Dis Relat Surg
2025;36(1):193-199

ORIGINAL ARTICLE

Evaluation of ChatGPT-4o's answers to questions about hip arthroscopy from the patient perspective

Gökhan Ayık, MD, Niyazi Ercan, MD, Yunus Demirtaş, MD, Tuğrul Yıldırım, MD, Gökhan Çakmak, MD

Department of Orthopedics and Traumatology, Yüksek İhtisas University, Ankara, Türkiye

ABSTRACT

Objectives: This study aimed to evaluate the responses provided by ChatGPT-4o to the most frequently asked questions by patients regarding hip arthroscopy.

Materials and methods: In this cross-sectional survey study, a new Google account without a search history was created to determine the 20 most frequently asked questions about hip arthroscopy via Google. These questions were asked to a new ChatGPT-4o account on June 1, 2024, and the responses were recorded. Ten orthopedic surgeons specializing in sports surgery rated the responses using a rating scale to assess relevance, accuracy, clarity, and completeness. The responses were scored on a scale from 1 to 5, with 1 being the worst and 5 being the best. The interrater reliability assessed via the intraclass correlation coefficient (ICC).

Results: The lowest score given by the surgeons for any response was 4/5 in each subcategory. The highest mean scores were in accuracy and clarity, followed by relevance, with completeness receiving the lowest scores. The overall mean score was 4.49±0.16. Interrater reliability showed insufficient overall agreement (ICC=0.004, p=0.383), with the highest agreement in clarity (ICC=0.039, p=0.131) and the lowest in accuracy (ICC=–0.019, p=0.688).

Conclusion: The study confirms our hypothesis that ChatGPT-4o provides above-average quality responses to frequently asked questions about hip arthroscopy, as evidenced by the high scores in relevance, accuracy, clarity, and completeness. However, it is still advisable to consult orthopedic specialists on the subject, incorporating ChatGPT's suggestions during the final decision-making process.

Keywords: Artificial intelligence, ChatGPT-4o, hip arthroscopy, patient education.

Femoroacetabular impingement (FAI) is one of the most prominent issues in contemporary orthopedics.[1] It occurs due to abnormal contact between the femur and acetabulum, each exhibiting abnormal morphological properties.[2] Recently, hip arthroscopy has advanced rapidly and become widely practiced. It is now a frequently performed surgery with various indications, particularly in FAI.[1,3-5]

The development of artificial intelligence (AI) has accelerated significantly over the past 20 years, affecting nearly every aspect of life, including medicine. Artificial intelligence refers to technologies or machines capable of performing tasks such as problem-solving, learning, language interpretation, pattern recognition, and planning.[6] Recently, easily accessible AI applications such as ChatGPT (OpenAI Inc., San Francisco, CA, USA) have become prevalent, generating human-like text, answering questions across various domains, and engaging in natural language conversations.[7] On May 13, 2024, OpenAI introduced the new ChatGPT-4o, marking a significant revolution in the field. The integration of AI into the fields of medicine and orthopedics appears inevitable under these circumstances. With the introduction of this latest version of ChatGPT,

Received: August 27, 2024
Accepted: December 11, 2024
Published online: December 18, 2024

Correspondence: Gökhan Ayık, MD. Yüksek İhtisas Üniversitesi Ortopedi ve Travmatoloji Anabilim Dalı, 06530 Çankaya, Ankara, Türkiye.

E-mail: drgayik@gmail.com

Citation: Ayık G, Ercan N, Demirtaş Y, Yıldırım T, Çakmak G. Evaluation of ChatGPT-4o's answers to questions about hip arthroscopy from the patient perspective. Jt Dis Relat Surg 2025;36(1):193-199. doi: 10.52312/jdrs.2025.1961.

©2025 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/).

many patients and individuals are increasingly using these tools to access medical information. However, there is currently a lack of clear data regarding the accuracy and reliability of the information provided by ChatGPT-4o.

This study aimed to evaluate the responses provided by ChatGPT-4o to the most frequently asked patient questions about hip arthroscopy. The evaluation will be based on relevance, accuracy, clarity, and completeness. We hypothesized that ChatGPT-4o would deliver responses above average in quality.

MATERIALS AND METHODS

In this cross-sectional survey study, a new Google (Alphabet Inc, Mountain View, CA, USA) account with no search history was created to search for "frequently asked questions about hip arthroscopy" on Google (www.google.com). Using the "other questions" section on the main screen, we identified a total of 100 initial questions. These questions were reviewed by two researchers with experience in hip arthroscopy who consolidated repeat or similar questions, ultimately narrowing the list down to 20 unique questions (Figure 1). To interact with ChatGPT-4o, we created a new, unused, paid ChatGPT-4o account that had not been used for any prior topics. These questions were posed to ChatGPT-4o on June 1, 2024, but not in immediate succession; instead, each question was asked at different times throughout the day to minimize potential bias from consecutive questioning. The researchers recorded the responses (Appendix), which were then used to create a survey. Ethical approval was not required for this study, as it involved only the analysis of an online tool without any human subject involvement.

In this study, we adopted a rating scale inspired by Magruder et al.'s[7] methodology for assessing large language models' responses to clinically relevant questions. Magruder et al.[7] evaluated ChatGPT's answers on six key characteristics: relevance, accuracy, clarity, completeness, evidence-based content, and consistency. Each characteristic was defined with specific criteria to guide evaluators, ensuring consistent and objective assessments.

Question number	Question
1	What movements and positions to avoid after hip arthroscopy?
2	What are the risks of hip arthroscopy?
3	How long does it take to fully recover from hip arthroscopy?
4	Can you walk after hip arthroscopy?
5	What happens if hip arthroscopy does not work?
6	How do you sit on the toilet after hip arthroscopy?
7	When can I start having sex after hip arthroscopy?
8	Can I drive after hip arthroscopy?
9	Can you walk up stairs after hip arthroscopy?
10	Do you always need a brace after hip arthroscopy?
11	When should I start physio after hip arthroscopy?
12	What is the success rate of hip arthroscopy?
13	What exercises can I do after hip arthroscopy?
14	What is the best sitting position after hip arthroscopy?
15	Can hip arthroscopy cause arthritis?
16	What is an alternative to hip arthroscopy?
17	How do you wear clothes after hip arthroscopy?
18	What are the do's and don'ts after hip arthroscopy?
19	What kind of anesthesia is used for hip arthroscopy?
20	What is the age limit for hip arthroscopy?

FIGURE 1. The 20 most frequently asked questions about hip arthroscopy, determined by authors, were posed to ChatGPT-4o.

TABLE I
Descriptive statistics of scores

	Number of Questions	Mean±SD
Relevance	20	4.49±0.17
Accuracy	20	4.51±0.15
Clarity	20	4.51±0.18
Completeness	20	4.46±0.15
Total	80	4.49±0.16

SD: Standard deviation.

In the rating scale used by Magruder et al.,[7] each criterion was scored on a scale from 1 to 5, with higher scores indicating better performance. For our study, we utilized four of these criteria, relevance, accuracy, clarity, and completeness, to maintain focus on aspects most critical to patient understanding in hip arthroscopy. Relevance included evaluating whether the answer directly addressed the question posed. Accuracy comprised determining if the information was correct. Clarity assessed the clarity and organization of the response, ensuring it was easy to understand. Completeness examined whether the answer covered all necessary information to fully respond to the question. The "consistency" criterion was excluded to mitigate potential variability in responses that may arise when identical or similar questions are posed repeatedly to the AI model, as each question was asked only once. Additionally, we did not include the "evidence-based content" criterion, as ChatGPT does not provide references for each response. Furthermore, due to the risk of AI hallucinations, the model may generate fictitious references, which could compromise the reliability of this criterion.[8] The modified scale thus retains the rigor of Magruder et al.'s[7] approach, while aligning with the specific needs of our study.

Ten orthopedic surgeons specializing in sports surgery, each with at least five years of experience, were asked to rate these responses using the rating scale applied by Magruder et al.[7] The relevance criterion was assessed by asking the question, "Is the provided answer directly related to the question asked?" The accuracy criterion was evaluated by asking, "Is the answer to the question correct?" The clarity criterion was assessed by asking, "Is the answer clear and understandable?" Finally, the completeness criterion was evaluated by asking, "Does the answer cover all aspects of the question and include all the necessary information to adequately address the question?" The surgeons were unaware that the responses were generated by ChatGPT-4o and evaluated them accordingly.

Statistical analysis

Data were analyzed using IBM SPSS version 23.0 software (IBM Corp., Armonk, NY, USA). Interrater reliability (IRR) was assessed using the intraclass correlation coefficient (ICC). The analysis results were presented as means and standard deviations.

RESULTS

Evaluation of the survey scores provided by 10 orthopedic sports surgeons revealed that the lowest score given for any response was four out of five in each subcategory of the rating scale, indicating consistently high-quality responses with minor areas for improvement. In general, ChatGPT's responses to each question frequently included recommendations

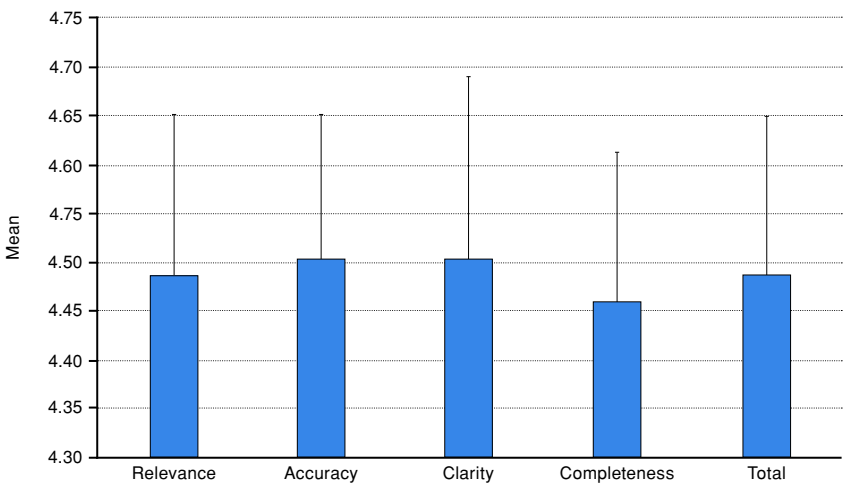


FIGURE 2. Graph of average scores and standard deviations.

TABLO II
Analysis of inter-rater reliability

	ICC	95% CI		p
		Lower bound	Upper bound	
Relevance	0.014	−0.038	0.13	0.313
Accuracy	−0.019	−0.058	0.072	0.688
Clarity	0.039	−0.023	0.171	0.131
Completeness	−0.008	−0.052	0.092	0.556
Total	0.004	−0.024	0.046	0.383

CI: Confidence interval; ICC: Intraclass correlation coefficient.

to consult with a surgeon or physical therapist at the end of the answer. Additionally, it often offered the option to seek further information. The responses were generally organized in a categorized manner and indicated that personalized recommendations might be necessary.

The highest mean scores were observed in accuracy and clarity, followed by relevance, with completeness receiving the lowest scores (Table I). Means and standard deviations are illustrated in Figure 2.

In assessing IRR, overall agreement among evaluators was insufficient (ICC=0.004; p=0.383). Clarity showed the highest agreement, while accuracy had the lowest (ICC=0.039, p=0.131 *vs.* ICC=-0.019, p=0.688; Table II). A bar chart displaying the ICC is shown in Figure 3.

To interpret the ICC values meaningfully, the ICCs must first reach statistical significance. However, the interrater agreement results indicated that the p-values did not reach the threshold for significance. Consequently, regardless of the ICC values, the lack of significant agreement renders them uninterpretable. Additionally, the negative ICC values observed may be attributed to low variance among raters. When there is insufficient variance across evaluators, the ICC can yield negative results. This indicates low reliability and a lack of consistent agreement among raters.

DISCUSSION

This study found that the overall scores for ChatGPT-4o's responses to the questions were high, with the lowest score being 4 on a 5-point scale. The mean scores for relevance, accuracy, clarity, and completeness were 4.49±0.17, 4.51±0.15, 4.51±0.18, and 4.46±0.15, respectively, confirming our hypothesis. Each response provided by ChatGPT-4o received above-average scores in relevance, accuracy, clarity, and completeness. However, IRR was poor, with the

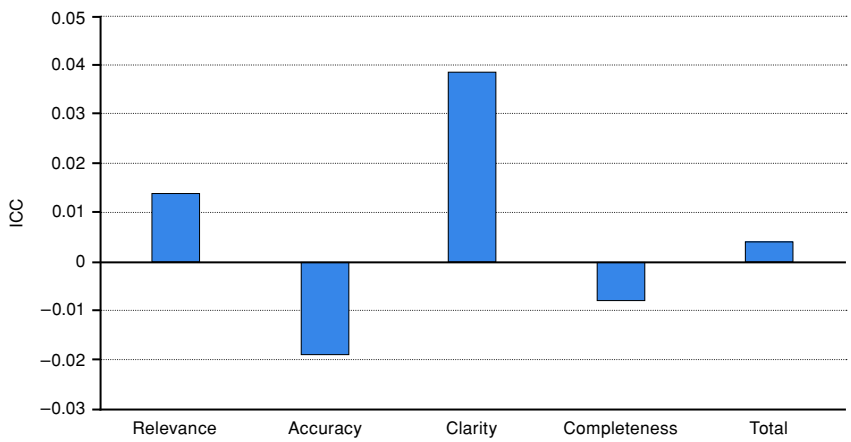


FIGURE 3. Bar chart of ICC coefficients.
ICC: Intraclass correlation coefficient.

lowest agreement in accuracy and the highest in clarity. In a study conducted by Magruder et al.[7] on total knee arthroplasty, the overall IRR was found to be poor. This study aimed to assess the quality of ChatGPT's responses to questions derived from the American Academy of Orthopedic Surgeons Clinical Practice Guidelines. These responses were evaluated by fellowship-trained surgeons. The study highlighted that while ChatGPT demonstrated above-average accuracy in answering questions, its reliability varied significantly.

The low IRR observed in our study may reflect challenges in achieving consistent agreement among evaluators. However, it is important to note that the variability in ratings primarily occurred between scores of 4 and 5. Such minor differences are unlikely to carry significant clinical implications, as they reflect only subtle variations in expert opinion. These slight discrepancies likely arise from the diverse perspectives and experiences of individual surgeons rather than a fundamental disagreement on the quality of the responses. Consequently, while the IRR values were lower than expected, this outcome is unlikely to undermine the validity of the findings in a meaningful way.

Despite its benefits, hip arthroscopy carries significant risks.[3] Clarke et al.[9] reported a complication rate of 1.4%, one of the first such data points in the literature. This rate has slightly increased as the number of procedures has grown, and relatively inexperienced surgeons have performed hip arthroscopy.[10] Currently, informing patients about potential complications is crucial. Additionally, patients often seek detailed information about this procedure. This trend highlights the need for patients to research and seek reliable information. The internet is a vital tool for this purpose, with Google already playing a significant role.[11,12] However, in the future, AI may take over this role, as it is transforming medicine.[13]

Artificial intelligence traditionally refers to the theory that computers can learn to perform tasks through pattern recognition with minimal human involvement. A more modern and accurate definition of AI is the application of algorithms that enable machines to solve problems traditionally requiring human intelligence.[6] ChatGPT is an AI-powered chatbot designed to respond to users' requests. Since its launch, ChatGPT has become a popular application, attracting millions of users in a short time.[14] As the data input for these chatbots increases, their capabilities improve, leading to new versions. The release of ChatGPT-4o marked a significant technological leap.

The volume of data worldwide is growing exponentially. It is estimated that medical data doubled every 50 years in the 1950s, every seven years in 1980, and every three and a half years in 2010, while today, it doubles every 73 days.[15] Artificial intelligence is increasingly used to classify, interpret, and make this massive data load accessible.[13] ChatGPT by OpenAI has found applications in many areas of life and has become a valuable tool for patients seeking medical information.[7] ChatGPT can process vast amounts of data, generate content, access information, and translate it into the desired language.

Although orthopedics has lagged in adopting AI, the integration process is accelerating.[13] The widespread use of these chatbots has increased their academic utility and allowed patients to use them to access medical information.[14] Gilson et al.[16] demonstrated that ChatGPT could answer USMLE (United States Medical Licensing Examination) questions at the level of nearly a third-year medical student. Furthermore, many studies are being conducted in various fields of medicine based on the answers provided by these chatbots.[7] These studies include shoulder stabilization procedures,[17] hip and knee arthroplasties,[7,14] anterior cruciate ligament reconstruction,[18] and more. While some studies found the answers satisfactory, others found them lacking. However, to the best of our knowledge, no study has evaluated the responses of this version of ChatGPT (ChatGPT-4o) to patient questions regarding hip arthroscopy.

AlShehri et al.'s[19] study evaluated the educational potential of ChatGPT version 3.5 in answering common patient questions about hip arthroscopy, grading responses based on accuracy and completeness. Their study utilized a four-grade system (A-D) and highlighted that while ChatGPT could provide satisfactory answers, inaccuracies were present, warranting caution in its use for patient education. They found that eight out of 10 responses were rated "B" or higher on a 4-point grading scale (A being the best and D the worst), although one answer was incorrect. In contrast, our study uses ChatGPT-4o, an updated version, with a focus on specific dimensions of response quality (relevance, accuracy, clarity, and completeness) to provide a more nuanced evaluation of the AI's performance. Furthermore, we employed multiple raters to assess IRR, which revealed significant variability in ratings, underscoring the subjectivity

inherent in assessing AI-generated content. Sparks et al.[20] reported that ChatGPT-3.5 performed reasonably well in providing general information about common orthopedic conditions but lacked details on risk factors and treatment options. This evaluation suggests that while ChatGPT holds potential as a patient education tool, its limitations must be carefully considered, particularly as different versions may vary in accuracy and reliability.

Özbek et al.[21] evaluated ChatGPT-4.0's responses to 25 common patient questions about hip arthroscopy, focusing on the accuracy of answers. Their study used a 4-point rating scale, where responses were rated from "excellent" to "unsatisfactory" based on the need for clarification. The results demonstrated that ChatGPT-4.0 provided primarily "excellent" responses, with only two questions requiring minimal clarification, indicating a high level of accuracy and reliability. Our study not only evaluated ChatGPT-4.0's accuracy but also assessed ChatGPT-4o and additional aspects of response quality (relevance, clarity, and completeness) providing a more comprehensive evaluation framework. Furthermore, our use of multiple raters highlighted interrater variability, a dimension not explored in Özbek et al.'s[21] study. While their findings support ChatGPT as a potential supplementary tool for patient education, our study, although also yielding promising results, adds a different perspective by suggesting that variability in ratings may impact the consistency of its educational value. It should not be overlooked that the scores were between 4/5 and 5/5; however, the variability in ratings must also be taken into consideration. The presence of negative values in ICCs may indicate insufficient variance among raters, suggesting low reliability and a lack of consistent agreement. Therefore, more comprehensive studies are needed to further evaluate the clinical applicability of ChatGPT-4o as a tool for patient education.

It remains uncertain whether such chatbots can consistently define, express, and convey accurate information. Therefore, the information they provide should be critically evaluated.[13] While humans are currently needed to manage AI, this may change in the future as the amount and accuracy of data increase. The key question is whether AI and the information it provides can be trusted. In our study, ChatGPT-4o generally provided satisfactory answers regarding hip arthroscopy. However, due to the lack of a clear consensus on the pre- and postmanagement of hip arthroscopy and the potential for variation among surgeons, it is prudent to view these results with some skepticism. Moreover, IRR was poor in this study.

In clinical practice, patients can use ChatGPT-4o to obtain information about hip arthroscopy. However, for final decisions and outcomes, consulting an orthopedic surgeon is strongly recommended. While it is possible to predict the future trajectory of AI development, it is also important to acknowledge that advancements may surpass current expectations. Just as the emergence of AI in recent years has had a revolutionary impact, its rapid development continues at an unprecedented pace. Data obtained from studies such as this one can contribute to the refinement of AI systems, potentially paving the way for the creation of personalized applications that patients may use in the future.

This study had several limitations. First, the number of evaluators was limited to 10, and the evaluation process was inherently subjective. Second, AI applications such as ChatGPT are based on machine learning and may produce different responses at different times. To address this variability, we asked the questions at different times using a new account. Additionally, there is currently no standardized system for scoring AI-generated responses. Another limitation was that although the scores ranged between 4/5 and 5/5, their interpretation may differ statistically and clinically. Furthermore, this study utilized ChatGPT-4o, a paid version of the AI model, while similar studies often use the free version. As with any new technology, initial applications tend to incur higher costs, but these typically decrease over time. Thus, while this limitation is relevant now, it may become less significant in the future.

In conclusion, this study evaluated ChatGPT-4o's responses to frequently asked patient questions about hip arthroscopy, focusing on relevance, accuracy, clarity, and completeness. Orthopedic surgeons rated the responses, yielding an overall high mean score, with the highest scores in accuracy and clarity. Despite poor interrater reliability, which highlights variability in response quality perception, ChatGPT-4o demonstrates significant potential as a supplementary patient education tool. With continued advancements and careful integration into clinical practice, it could serve as a valuable adjunct in improving patient understanding of medical procedures. However, it is essential to emphasize that ChatGPT-4o should not replace professional medical advice, and patients are strongly encouraged to consult orthopedic specialists to confirm AI-provided information.

Data Sharing Statement: The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Author Contributions: All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed: G.A., N.E., T.Y. The first draft of the manuscript was written by G.A. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

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. Çiçeklidağ M, Ayanoğlu T, Kaptan AY, Vural A, Kalaycıoğlu O, Özer M, et al. Effect of the presence of cysts in the hip joint on hip arthroscopy. Jt Dis Relat Surg 2024;35:645-53. doi: 10.52312/jdrs.2024.1657.
2. Ganz R, Parvizi J, Beck M, Leunig M, Nötzli H, Siebenrock KA. Femoroacetabular impingement: A cause for osteoarthritis of the hip. Clin Orthop Relat Res 2003;:112-20. doi: 10.1097/01.blo.0000096804.78689.c2.
3. Jamil M, Dandachli W, Noordin S, Witt J. Hip arthroscopy: Indications, outcomes and complications. Int J Surg 2018;54:341-4. doi: 10.1016/j.ijsu.2017.08.557.
4. Perets I, Rybalko D, Mu BH, Friedman A, Morgenstern DR, Domb BG. Hip arthroscopy: Extra-articular procedures. Hip Int 2019;29:346-54. doi: 10.1177/1120700019840729.
5. Divecha HM, Rajpura A, Board TN. Hip arthroscopy: A focus on the future. Hip Int 2015;25:323-9. doi: 10.5301/hipint.5000271.
6. Atik OŞ. Artificial intelligence: Who must have autonomy the machine or the human? Jt Dis Relat Surg 2024;35:1-2. doi: 10.52312/jdrs.2023.57918.
7. Magruder ML, Rodriguez AN, Wong JCJ, Erez O, Piuzzi NS, Scuderi GR, et al. Assessing ability for ChatGPT to answer total knee arthroplasty-related questions. J Arthroplasty 2024;39:2022-7. doi: 10.1016/j.arth.2024.02.023.
8. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: Implications in scientific writing. Cureus 2023;15:e35179. doi: 10.7759/cureus.35179.
9. Clarke MT, Arora A, Villar RN. Hip arthroscopy: complications in 1054 cases. Clin Orthop Relat Res 2003;406:84-8. doi: 10.1097/01.blo.0000043048.84315.af.
10. Harris JD, McCormick FM, Abrams GD, Gupta AK, Ellis TJ, Bach BR Jr, et al. Complications and reoperations during and after hip arthroscopy: A systematic review of 92 studies and more than 6,000 patients. Arthroscopy 2013;29:589-95. doi: 10.1016/j.arthro.2012.11.003.
11. Cocco AM, Zordan R, Taylor DM, Weiland TJ, Dilley SJ, Kant J, et al. Dr Google in the ED: Searching for online health information by adult emergency department patients. Med J Aust 2018;209:342-7. doi: 10.5694/mja17.00889.
12. Van Riel N, Auwerx K, Debbaut P, Van Hees S, Schoenmakers B. The effect of Dr Google on doctor-patient encounters in primary care: A quantitative, observational, cross-sectional study. BJGP Open 2017;1:bjgpopen17X100833. doi: 10.3399/bjgpopen17X100833.
13. Kunze KN, Orr M, Krebs V, Bhandari M, Piuzzi NS. Potential benefits, unintended consequences, and future roles of artificial intelligence in orthopaedic surgery research : A call to emphasize data quality and indications. Bone Jt Open 2022;3:93-7. doi: 10.1302/2633-1462.31.BJO-2021-0123.R1.
14. 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;35:169-76. doi: 10.52312/jdrs.2023.1402.
15. Densen P. Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc 2011;122:48-58.
16. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312. doi: 10.2196/45312.
17. Hurley ET, Crook BS, Lorentz SG, Danilkowicz RM, Lau BC, Taylor DC, et al. Evaluation high-quality of ınformation from ChatGPT (Artificial Intelligence-Large Language Model) artificial intelligence on shoulder stabilization surgery. Arthroscopy 2024;40:726-31.e6. doi: 10.1016/j.arthro.2023.07.048.
18. Johns WL, Martinazzi BJ, Miltenberg B, Nam HH, Hammoud S. ChatGPT provides unsatisfactory responses to frequently asked questions regarding anterior cruciate ligament reconstruction. Arthroscopy 2024;40:2067-79.e1. doi: 10.1016/j.arthro.2024.01.017.
19. AlShehri Y, McConkey M, Lodhia P. ChatGPT provides satisfactory but occasionally inaccurate answers to common patient hip arthroscopy questions. Arthroscopy 2024:S0749-806300452-3. doi: 10.1016/j.arthro.2024.06.017.
20. Sparks CA, Fasulo SM, Windsor JT, Bankauskas V, Contrada EV, Kraeutler MJ, et al. ChatGPT is moderately accurate in providing a general overview of orthopaedic conditions. JB JS Open Access 2024;9:e23.00129. doi: 10.2106/JBJS.OA.23.00129.
21. Özbek EA, Ertan MB, Kından P, Karaca MO, Gürsoy S, Chahla J. ChatGPT can offer at least satisfactory responses to common patient questions regarding hip arthroscopy. Arthroscopy 2024:S0749-806300640-6. doi: 10.1016/j.arthro.2024.08.036.