Joint Diseases and Related Surgery

Jt Dis Relat Surg
2026;37(1):i-xiv

ORIGINAL ARTICLE

Comparison of ChatGPT and Google in addressing patients' questions on robot-assisted total hip arthroplasty

Mustafa Fatih Dasci, MD[1], Serkan Surucu, MD[2], Furkan Aral, MD[3], Mahmud Aydin, MD[4], Cihangir Turemis, MD[5], N Amir Sandiford, MD[6], Mustafa Citak, MD[7]

[1]Department of Orthopedics and Traumatology, University of Health Science, Bağcılar Training and Research Hospital, İstanbul, Türkiye
[2]Department of Orthopaedics and Rehabilitation, Yale University, New Haven, USA
[3]Department of Orthopedics and Traumatology, Gazi University Faculty of Medicine, Ankara, Türkiye
[4]Department of Orthopedics and Traumatology, Memorial Şişli Hospital, İstanbul, Türkiye
[5]Department of Orthopedics and Traumatology, Çeşme Alper Çizgenat State Hospital, İzmir, Türkiye
[6]Joint Reconstruction Unit, Southland Hospital, University of Otago, Invercargill, New Zealand
[7]Department of Orthopaedic Surgery, HELIOS ENDO-Klinik Hamburg, Hamburg, Germany

Total hip arthroplasty (THA) remains the definitive treatment modality for conditions requiring hip replacement, such as end-stage hip joint osteoarthritis and hip fractures.[1,2] Robot-assisted THA (RATHA) constitutes a substantial advancement in orthopedic surgery, offering enhanced precision and potential improvements in patient outcomes.[3,4] Advancements in artificial intelligence (AI), machine learning, and robotics have facilitated the integration of robotic systems into orthopedic procedures over the past decade. These technologies enable surgeons to perform hip replacement surgeries with enhanced accuracy, potentially leading to

Received: May 13, 2025
Accepted: July 27, 2025
Published online: October 31, 2025

Correspondence: Mustafa Citak, MD, PhD. ENDO-Klinik Hamburg, Department of Orthopaedic Surgery, Holstenstr. 2, 22767 Hamburg, Germany.

E-mail: mcitak@gmx.de

Doi: 10.52312/jdrs.2026.2368

Citation: Dasci MF, Surucu S, Aral F, Aydin M, Turemis C, Sandiford NA, et al. Comparison of ChatGPT and Google in addressing patients' questions on robot-assisted total hip arthroplasty. Jt Dis Relat Surg 2026;37(1):i-xiv. doi: 10.52312/jdrs.2026.2368.

©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/).

ABSTRACT

Objectives: This study aims to compare ChatGPT (Generative Pre-Trained Transformer) and Google in addressing frequently asked questions (FAQs), answers, and online sources regarding robot-assisted total hip arthroplasty (RATHA).

Materials and methods: On December 15th, 2024, the 20 most FAQs were identified by inputting the search term "Robot-Assisted Total Hip Replacement" into both Google Search and ChatGPT-4o. Twenty FAQs were independently identified using a clean Google search and a prompt to ChatGPT-4o. The FAQs on Google were sourced from the "People also ask" section, while ChatGPT was requested to generate the 20 most often asked questions. All questions, answers, and references cited were recorded. A modified version of the Rothwell system was used to categorize questions into 10 subtopics: special activities, timeline of recovery, restrictions, technical details, cost, indications/management, risks and complications, pain, longevity, and evaluation of surgery. Each reference was categorized into the following groups: commercial, academic, medical practice, single surgeon personal, or social media. Responses were also graded as "excellent response not requiring clarification" (1), "satisfactory requiring minimal clarification" (2), "satisfactory requiring moderate clarification" (3), or "unsatisfactory requiring substantial clarification" (4).

Results: Overall, 20% of the questions that Google and ChatGPT-4o considered as the most FAQ were similar to each other. Technical details (35%) were the most common categories of questions. The ChatGPT provided significantly more academic references than Google search (70% *vs*. 20%, p=0.0113). Conversely, Google web search cited more medical practice references (40% *vs*. 0%, p=0.0033), single surgeon websites (20% *vs*. 0%, p=0.1060), and government websites (10% *vs*. 0%, p=0.4872) more frequently than ChatGPT. In terms of response quality, 62% of answers were rated as Grade 1-2 (excellent or satisfactory with minimal clarification), while 38% required moderate or substantial clarification (Grades 3-4).

Conclusion: ChatGPT demonstrated comparable results to those of Google searches on information regarding RATHA, with a higher reliance on academic sources. While most responses were satisfactory, a notable proportion required further clarification, emphasizing the need for continued evaluation of these platforms to ensure accuracy and reliability in patient education. Taken together, these technologies have the capacity to enhance health literacy and provide enhanced shared decision-making for patients seeking information on RATHA.

Keywords: Artificial intelligence, ChatGPT, clinical relevance, Google, health information quality, robot-assisted total hip arthroplasty, patient education.

improved restoration of the center of rotation, joint biomechanics, reduced tissue damage, and faster recovery.[5] Traditional approaches, while effective, carry the inherent risks of malalignment or component malpositioning, which may affect long-term outcomes. Robotic assistance addresses these limitations by offering real-time intraoperative guidance, three-dimensional (3D) imaging, and pre-surgical planning capabilities. The precision and reproducibility of RATHA have sparked growing interest in the medical community, with increasing adoption across hospitals worldwide. However, well-designed, prospective, controlled trials with long-term follow-up are still warranted to evaluate the efficacy of RATHA.[6]

Moreover, the incorporation of AI tools, such as ChatGPT, into medical decision-making and patient education has opened new avenues for improving healthcare delivery. Of note, AI-powered platforms have been used to provide patients with accessible, tailored information regarding RATHA, offering responses to frequently asked questions (FAQs) in real-time. These platforms complement traditional search engines, such as Google, by offering more structured, conversational answers to patients' concerns. As a result, they are more intuitive for patients.

Unlike traditional search engines such as Google, which function by retrieving and listing hyperlinks to various external sources, AI-powered applications such as ChatGPT synthesize and generate complete responses directly. While the user plays an active role in reviewing, interpreting, and validating information retrieved via search engines, AI tools assume responsibility for content synthesis and summarization.[7] This autonomy in processing and presenting information distinguishes AI chatbots as interactive, answer-generating platforms, an essential conceptual distinction that frames the basis of this comparative study.

Recent studies have explored the efficacy of traditional search engines and AI-driven chatbots in disseminating medical knowledge.[8-10] Such studies are crucial for understanding how patients seek and process information regarding novel surgical techniques such as RATHA. In the present study, we aimed to compare the quality, accuracy, and relevance of answers, and online sources provided by Google and ChatGPT-4o to the most FAQs about RATHA and to examine how evolving AI technologies are transforming patient education into modern surgical practices.

MATERIALS AND METHODS

This single-center, cross-sectional study was conducted at Endo-Klinik Hamburg, Department of Orthopedic Surgery, on December 15th, 2024. The quality and clinical relevance of responses provided by ChatGPT-4o (OpenAI; San Francisco, CA) and Google (Menlo Park, CA, USA) to FAQs about RATHA were evaluated.

Methods were adapted from a previous study by Dubin et al.[11] The Google search engine was performed using a clean-installed Google Chrome browser (version 112.0.5615.137) with cleared cache and history to avoid personalized results. On December 15th, 2024, Google search was conducted using the key phrase: "robot-assisted total hip replacement". In Google searches, the questions were extracted from the 'People also ask' section, which displays commonly asked questions along with additional related questions that appear when each query is expanded. From the search, the top-listed questions were recorded. Questions from this section were included if the question included the term "robotic total hip replacement," and "robot-assisted total hip arthroplasty". Duplicate and irrelevant entries were excluded to form a final list of 20 FAQs. Google itself does not generate answers but displays a curated list of websites. We considered the top-ranked, featured excerpt or snippet as the "Google answer" when available. If no featured snippet existed, we selected the first website listed in the search results and used the main explanatory paragraph from that page as the representative response. The questions, answers, and online sources were recorded.

A new and clean ChatGPT-4o account was used to interact with the platform. The following statements were entered into ChatGPT-4o: "Perform a Google search with the search term robot-assisted total hip replacement and record the 20 most FAQs related to the search term with answers to the questions and the online source." The top 20 questions, answers, and sources provided by ChatGPT-4o were recorded (Table I) and screened for duplicates, unrelated and hallucinated content, resulting in a matched set of 20 FAQs from ChatGPT for comparison.

Data classification and evaluation

Following question identification, all responses were analyzed. The Rothwell's classification is primarily designed to understand the questions asked in a group.[12] Rothwell classification categorizes the questions and the online sources of the answers received according to their content.

TABLE I

Rothwell's Classification System for Questions and Websites

Robotic assisted total hip arthroplasty **(Google)**	Grade	Robotic assisted total hip arthroplasty **(ChatGPT)**	Grade
What are the disadvantages of robotic hip replacement? (Technical Details) The disadvantages of robotic-assisted joint replacement surgery include: Possible need for additional imaging (CT scan) before surgery. Possible risk of bone fractures where the navigation pins are placed. **(Academic)**	2	**What is Robotic Assisted Total Hip Arthroplasty? (Indications/Management)** Robotic Assisted THA uses robotic systems to aid surgeons in the precise placement and alignment of hip implants, enhancing the accuracy and outcomes of hip replacement surgeries **(Academic)**	2
What is robotic arthroplasty total hip replacement? (Technical Details) A robotic hip replacement is similar to a traditional hip replacement. Your surgeon removes damaged tissue in your hip and replaces it with an artificial joint. The difference is that it's done with assistance from a robotic arm, which allows for greater precision. **(Academic)**	2	**How does robotic technology improve hip replacement surgery? (Technical Details)** Robotic technology provides a 3D model of the patient's hip, allows for precise planning, and aids in the accurate placement of implants, reducing the risk of complications and improving recovery times **(Academic)**	2
***How long does it take to recover from robotic hip replacement? (Timeline of recovery)** The average recovery period for this procedure is three to six months. This is because no two people recover in exactly the same manner. Some patients can do most of their usual daily activities as early as three months post-surgery. Others, however, take a longer time to heal. **(Government)**	2	**Who is a candidate for Robotic Assisted THA? (Indications/Management)** Candidates typically suffer from degenerative joint diseases such as osteoarthritis, rheumatoid arthritis, or hip dysplasia, and have not responded to conservative treatments **(Academic)**	1
Is robot-assisted hip replacement better? (Indications/Management) Robot-assisted procedures help ensure that your prosthetic joint fits perfectly. Other benefits include smaller incisions, faster recovery times, and better results than traditional open surgery. **(Single Surgeon)**	4	**What are the benefits of Robotic Assisted THA? (Technical Details)** Benefits include more accurate implant placement, reduced risk of dislocation, improved leg length uniformity, quicker recovery times, and potentially longer implant lifespan **(Academic)**	2
Who should not have robotic surgery? (Indications/Management) These are situations that would result in a high risk of harm. For example, robotic surgery should not be performed in a critically ill patient who needs emergency surgery or a trauma patient. Some types of comorbidities. Comorbidity is having more than one medical condition at a time. **(Medical Practice)**	3	**What are the risks associated with Robotic Assisted THA? (Risk)** While generally safe, risks include those common to any surgical procedure such as infection, blood clots, and implant issues. Specific robotic surgery risks are minimal but can include technical failures **(Academic)**	1
What can you never do again after a robotic hip replacement? (Timeline of recovery) Don't cross your legs at the knees for at least 6 to 8 weeks. Don't bring your knee up higher than your hip. Don't lean forward while sitting or as you sit down. Don't try to pick up something on the floor while you are sitting. Don't turn your feet excessively inward or outward when you bend down **(Medical Practice)**	1	**How is a 3D model used in Robotic Assisted THA? (Technical Details)** A CT scan creates a detailed 3D model of the patient's hip, which is used to plan the surgery and guide the robotic arm during the procedure for precise implant placement **(Academic)**	1

TABLE I

Continued

Robotic assisted total hip arthroplasty **(Google)**	Grade	Robotic assisted total hip arthroplasty **(ChatGPT)**	Grade
How big is the incision for robotic hip replacement? **(Technical Details)** For this type of minimally invasive hip replacement, the surgeon uses a single incision that usually measures 3 to 6 inches. The length of the incision depends on the size of the patient and the difficulty of the procedure **(Medical Practice)**	1	**How long does a Robotic Assisted THA surgery take?** **(Technical Details)** The duration is similar to traditional hip replacement surgeries, typically taking around 1-2 hours depending on the complexity of the case **(Academic)**	2
Which method of robotic total hip replacement is the best? **(Technical Details)** The anterior approach is an easier surgery to recover from. There is less muscle damage so less healing time is required. Patients are home within a day, sometimes the same day or the next, from an anterior hip operation **(Single Surgeon)**	3	**What is the recovery time for Robotic Assisted THA?** **(Timeline of recovery)** Recovery time varies, but many patients can expect a faster recovery compared to traditional surgery, often resuming normal activities within 6-12 weeks. **(Academic)**	3
Is robotic hip replacement more expensive? **(Cost)** Robotic hip replacement surgeries might cost more than traditional methods, but it's important to think about the long-term advantages, like better results and fewer complications. **(Single Surgeon)**	3	**What should patients expect during recovery after robotic total hip replacement?** **(Timeline of Recovery)** Patients should expect gradual improvement in mobility, with initial use of assistive devices like walkers or crutches, and a focus on physical therapy to regain strength and flexibility **(Academic)**	2
How do you wipe your bottom after robotic hip replacement? **(Timeline of Recovery)** Again, you can ask your surgeon for specific tips, but oftentimes it's helpful to lean on your non-operated hip for support and to rest that same arm on a counter for support if possible. Use your free hand on the opposite side to wipe without twisting your upper body **(Medical Practice)**	2	**Is Robotic Assisted THA more expensive than traditional THA?** **(Cost)** The cost may be higher due to the advanced technology used, but many patients find the benefits, such as reduced recovery time and improved outcomes, to justify the cost. **(Commercial)**	3
Who is not a candidate for robotic surgery? **(indications/Management)** However, not everyone is a good candidate for robotic surgery. Some of the contraindications for robotic surgery include severe obesity, bowel obstruction, and an inability to tolerate general anesthesia **(Single Surgeon)**	3	**How long do the implants last in robotic total hip replacement?** **(Timeline of Recovery)** With precise placement, implants from robotic-assisted surgeries can last 15-20 years or longer, similar to or better than those from traditional methods. **(Academic)**	2
What are lifelong restrictions after robotic hip replacement? **(Restrictions)** Avoiding high-impact activities like running, jumping, or lifting heavy weights can lessen the risk of damaging the replacement. If you do participate in these activities, you may weaken or loosen the replacement, which could require corrective surgery **(Medical Practice)**	2	**Are there any specific pre-operative preparations for Robotic Assisted THA?** **(Specific Activities)** Pre-operative preparations include physical evaluations, imaging studies, and discussions about medical history and medications to ensure patient readiness. **(Commercial)**	3
How long does it take to walk normally after robotic hip replacement surgery? **(Specific Activities)** How long does it take to walk normally after hip surgery? This depends on your general health and fitness, as well as the condition of your joints and muscles. However, in most cases, you will be able to walk without a walking aid after 10 to 14 days. **(Commercial)**	2	**What types of implants are used in Robotic Assisted THA?** **(Timeline of Recovery)** Implants are typically made of metal, ceramic, and plastic components, chosen based on the patient's specific needs and anatomical considerations **(Commercial)**	2

TABLE I

Continued

Robotic assisted total hip arthroplasty **(Google)**	Grade	Robotic assisted total hip arthroplasty **(ChatGPT)**	Grade
Does insurance cover robotic hip replacement? (Cost) Is there any additional out of pocket cost if I decided to have a robotic hip replacement? All insurance that covers traditional hip replacement cover the robotic hip replacement. There should be no additional cost incurred to the patient for using robotics during hip replacement surgery. **(Academic)**	4	**How does Robotic Assisted THA differ from traditional hip replacement? (Technical Details)** The key difference is the use of robotic technology for enhanced precision in planning and execution, which can lead to better outcomes and fewer complications **(Academic)**	3
Is robotic hip replacement covered by Medicare? (Cost) Medicare will cover almost any procedure that is deemed medically necessary. This includes robotic surgery. **(Commercial)**	2	**Can Robotic Assisted THA be performed on both hips simultaneously? (Technical Details)** Bilateral hip replacements can be performed, but the decision depends on the patient's overall health and specific circumstances. It is typically done in staged procedures. **(Commercial)**	1
What is a drawback of robotic surgery? (Technical Details) Robot-assisted surgery is only available in centers that have specially trained surgeons. Other disadvantages include: Complications that would require your surgeon to "switch" to an open procedure with larger incisions. **(Academic)**	3	**What kind of anesthesia is used during Robotic Assisted THA? (Technical Details)** General anesthesia or regional anesthesia (spinal or epidural) is commonly used, based on patient health, preferences, and surgeon recommendations. **(Academic)**	1
Do surgeons prefer robotic surgery? (Indications/management) Surgeons today often choose the robotic approach if it appears to have benefits for the patient compared to the conventional approach. **(Government)**	4	**How soon can I drive after robotic total hip replacement? (Specific Activities)** Patients usually return to driving within 2-6 weeks, depending on the side operated and use of narcotics. Right-sided surgery often requires a longer delay. **(Academic)**	2
What are the criticism of robotic surgery? (indications/Management) Critics frequently cite the longer operating times and higher costs of robotic surgery compared to conventional surgery. **(Medical Practice)**	2	**How do surgeons get trained in Robotic Assisted THA? (Technical Details)** Surgeons undergo specialized training and certification to use robotic systems, ensuring they are proficient in both the technology and surgical techniques required **(Academic)**	1
Can you ever kneel again after a robotic hip replacement? (Restrictions) The safe way to do this is to perform a single-legged kneel whereby the patient kneels on the knee of the operated side only. **(Medical Practice)**	2	**What are the common post-operative complications again after robotic total hip replacement? (Complications)** Complications can include infection, blood clots, hip dislocation, and implant issues, though these risks are minimized with robotic precision **(Commercial)**	4
How long does robotic hip surgery take? (Technical Detail) Your surgeon makes a small incision above your hip and removes the damaged parts of the joint. They then insert the artificial hip joints. Surgery takes about two to four hours. **(Medical Practice)**	1	**Can patients with metal allergies undergo Robotic Assisted THA? (Risks/Complications)** Yes, patients with metal allergies can have implants made from alternative materials like ceramics to avoid allergic reactions. **(Commercial)**	3

CT: Computed tomography; THA: Total hip arthroplasty.

The recorded questions were grouped under 10 subheadings according to modification of the Rothwell system (Table II).[11,12] Subheadings were as follows: Indications/Management, Technical Details, Evaluation of Surgery, Risks/Complications, Limitations, Special Activities, Recovery Timeline, Pain, Longevity, and Cost. The reported references for the answers to the most FAQs provided by each modality were categorized into the following groups: commercial, academic, medical practice, single surgeon personal, government, or social media.[11]

Once all responses were collected, two authors independently evaluated and graded them using the scoring system proposed by Mika et al.[13] Each response was given a numerical 'response accuracy score' based on its adequacy and the level of clarification required. Scores were categorized as: *(i)* excellent, requiring no clarification; *(ii)* satisfactory with minimal clarification; *(iii)* satisfactory with moderate clarification; or *(iv)* unsatisfactory, requiring substantial clarification (Table III).

Evaluation of data was performed by two independent reviewers. Any discrepancies in classification were resolved by consensus with a third reviewer. All reviewers are board certified orthopedic surgeons. These reviewers were blinded to whether the answer was from Google or ChatGPT-4o, as well as the source of information of the search engine.

Statistical analysis

Statistical analysis was performed using the IBM SPSS version 28.0 software (IBM Corp., Armonk, NY, USA). Continuous data were expressed in mean ± standard deviation (SD) or median (min-max), while categorical data were expressed in number and frequency. Cohen's kappa (κ) coefficients were determined to assess interobserver reliability. The κ value indicates the level of agreement among the observers. Landis and Koch classified κ values as follows: 0.00-0.20 indicates slight agreement; 0.21 to 0.40 denotes fair agreement; 0.41 to 0.60 reflects moderate agreement; 0.61 to 0.80 signifies substantial agreement; and values of 0.81 or greater

TABLE II

Comparison of Google and ChatGPT Questions and Responses

Rothwell's classification	**Description**
Fact	Asks whether something is true and to what extent, objective information
Policy	Asks whether a specific course of action should be taken to solve a problem
Value	Asks for evaluation of an idea, object, or event
Question classification by topic	**Description**
Fact	
Specific activities/restrictions	Ability/Inability to perform a specific activity after surgery
Cost	Cost of surgery including questions about insurance coverage
Timeline of recovery	Specific questions regarding recovery and timelines
Technical details	Surgical procedure, includes specific questions about surgery and anesthesia
Policy	
Indications	Surgical indications, alternatives, timing of surgery
Risks/complications	Risks/complications during and after surgery
Value	
Pain	Includes duration, severity, and management of pain
Evaluation of surgery	Evaluation of surgery, successfulness or invasiveness
Website categorization	**Description**
Commercial	Commercial public organization that provides source of health information,
Academic	Institutions including universities, academic medical centers, academic societies, and journals.
Medical practice	Local hospital or medical practice without an academic affiliation
Single surgeon	Websites of an individual surgeon
Government	Websites maintained by a national government organization.

TABLE III
Response rating system

Response Accuracy Score	Response Accuracy Description
1	Excellent response not requiring clarification
2	Satisfactory requiring minimal clarification
3	Satisfactory requiring moderate clarification
4	Unsatisfactory requiring substantial clarification

represent almost perfect agreement. The κ value for interobserver reliability was 0.90, indicating excellent agreement for website classification. The Fisher exact test for proportions was conducted to analyze question categories in relation to website classifications. A p value of <0.05 was considered statistically significant.

RESULTS

A total of 40 FAQs (20 from Google and 20 from ChatGPT) regarding RATHA were identified and analyzed. Subcategories of the most common FAQs are shown in Table IV.

Overall, 20% of the questions that Google and ChatGPT-4o considered as the most FAQ were similar to each other. According to the Rothwell classification, most questions fell into the Fact category for both platforms (ChatGPT: 70%, Google: 65%). The subcategories of the most FAQs based on their content are presented in Table IV. Overall, technical details (35%) were the most frequently addressed topic according to Rothwell's system. The most common subcategories by topic for ChatGPT-4o were technical details (40%) and timeline of recovery (20%) and risks/complications (15%); for Google web search, the most common subcategories were technical details (30%), indications/management (25%), timeline of recovery (15%) and cost (15%) (Table IV). Neither platform included questions about pain or evaluation of surgery. The categories of questions cited by Google

TABLE IV
Google and ChatGPT Questions Ratio

	Google (n=20)	ChatGPT (n=20)	Total (n=40)	
Category	n	n	n	p
Indications/management	5	2	7	0.4075
Technical details	6	8	14	0.7411
Evaluation of surgery	0	0	0	1.0
Risks/complications	0	3	3	0.2308
Restrictions	2	0	2	0.4872
Specific activities	1	2	3	1.0
Timeline of recovery	3	4	7	1.0
Pain	0	0	0	1.0
Longevity	0	0	0	1.0
Cost	3	1	4	0.6050
Total	20	20	40	
Website categories				
Academic	4	14	18	**0.0113**
Medical practice	8	0	8	**0.0033**
Single surgeon	4	0	4	0.1060
Government	2	0	2	0.4872
Commercial	2	6	8	0.2351
Total websites	20	20	40	

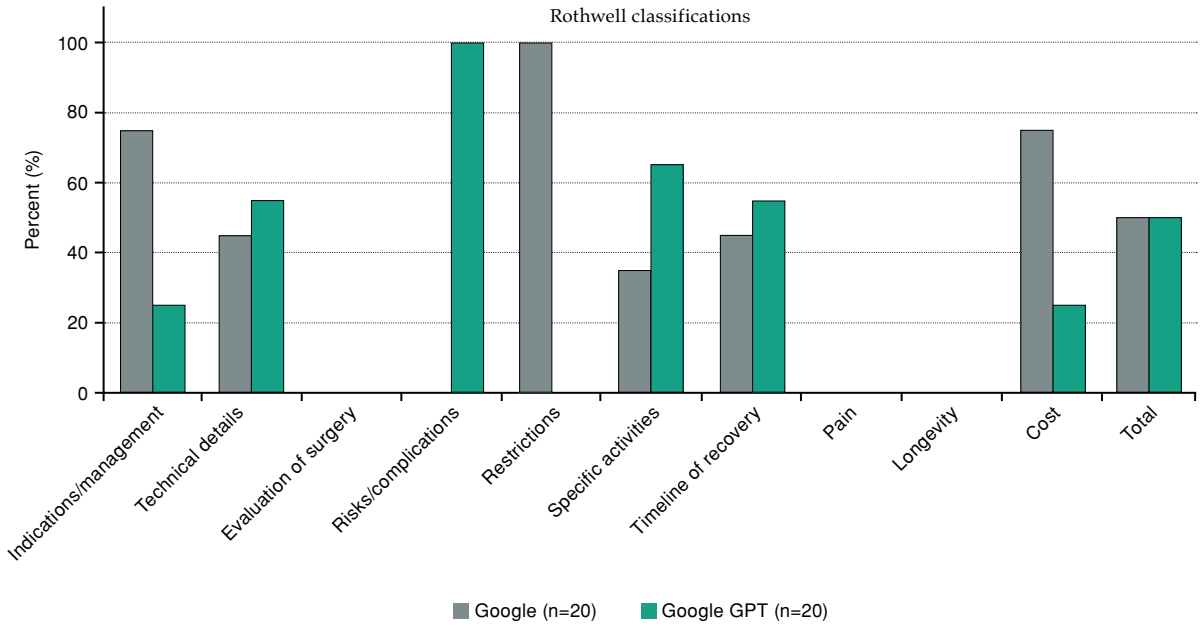

FIGURE 1. Rothwell classification and subgroups.

and ChatGPT included indications/management (25% *vs.* 10%, p=0.4075), technical details (30% *vs.* 40%, p=0.7411), surgical evaluation (0% *vs.* 0%), risks/complications (0% *vs.* 15%, p=0.2308), restrictions (10% *vs.* 0%, p=0.4872) and cost (15% *vs.* 5%, p=0.6050) (Figure 1). ChatGPT included more questions related to risk/complications (15%) compared to Google (0%), although this difference was not statistically significant (p=0.2308). The κ value for interobserver reliability was 0.95 (excellent agreement) for Rothwell's website classification system.

The distribution of information sources varied significantly between platforms (Figure 2). The most common sources of responses were medical practice (40%) on Google and academic (70%) on ChatGPT. The ChatGPT provided significantly more academic references than Google search (70% *vs.* 20%, p=0.0113) and this difference in source types was statistically

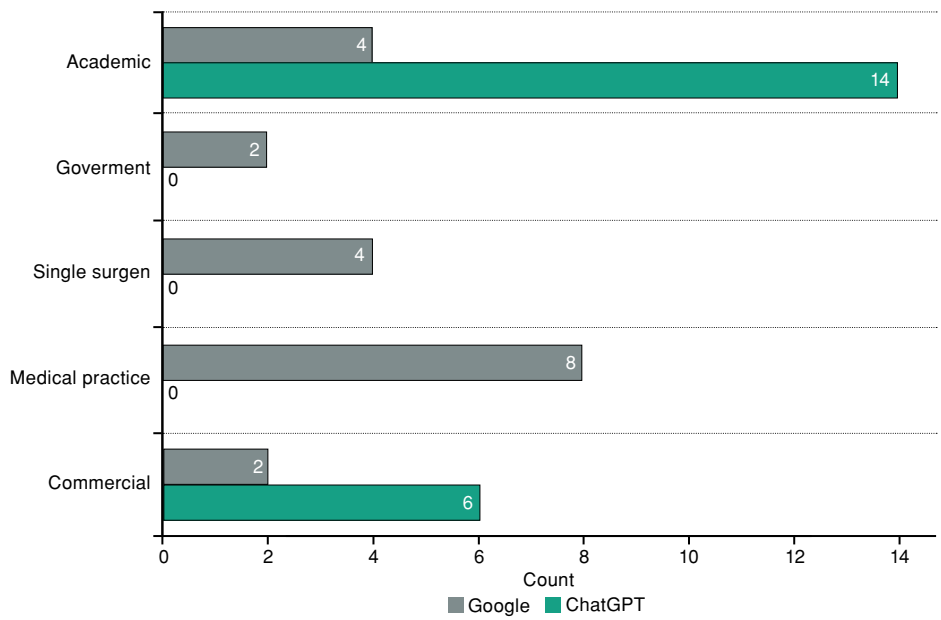

FIGURE 2. Website categories on ChatGPT and Google searching.

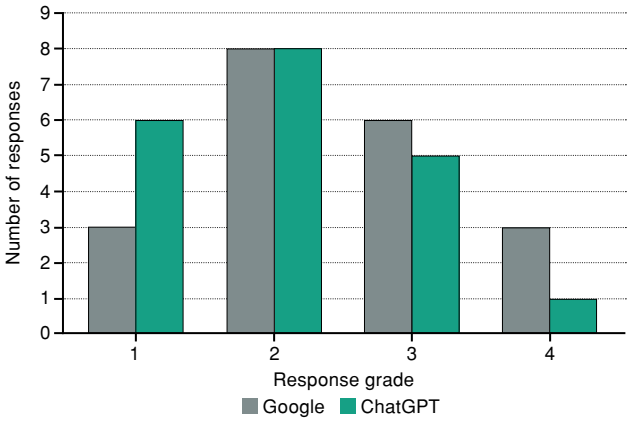

FIGURE 3. Grading of Google and ChatGPT responses to the most FAQs. Grades were defined as: (1) excellent response not requiring clarification, (2) satisfactory requiring minimal clarification, (3) satisfactory requiring moderate clarification, and (4) unsatisfactory requiring substantial clarification.

significant. In contrast, medical practice (40% *vs.* 0%, p=0.0033), single surgeon (20% *vs.* 0%, p=0.1060), and government (10% *vs.* 0%, p=0.4872) were cited more frequently by Google searches compared with ChatGPT. Source distribution is illustrated in Table IV and Figure 2.

All responses were collected, evaluated based on the response accuracy score,[13] and graded accordingly; these results are presented in Figure 3. A total of 40 FAQ responses from Google and ChatGPT were evaluated. Among these, nine (22%) were classified as Grade 1, 16 (40%) as Grade 2, 11 (27%) as Grade 3, and four (10%) as Grade 4. The grading of responses from Google and ChatGPT-4o was compared using the Wilcoxon signed-rank test, as the data were ordinal and not normally distributed ($p<0.05$). Google had a mean grading score of 2.45 (median 2.0), while ChatGPT had a mean of 2.05 (median 2.0), with lower scores indicating higher grade. Although ChatGPT demonstrated a trend toward greater grade, the difference was not statistically significant (p=0.190).

The chi-square goodness-of-fit analysis revealed no significant difference in the distribution of grades within each platform (Google: p=0.308; ChatGPT: p=0.158). These findings indicate that, within each platform, the frequency of Grades 1, 2, 3, and 4 was relatively balanced, with no single grade dominating the distribution.

DISCUSSION

In the present study, we aimed to compare the quality, accuracy, and relevance of answers, and online sources provided by Google and ChatGPT-4o to the most FAQs about RATHA. The main findings of this study were as follows: *(i)* Google web search and ChatGPT-4o produced extremely different results regarding the most FAQs and answers regarding RATHA, with minimal overlap in the questions; *(ii)* ChatGPT-4o provided a high percentage of academic sources, whereas Google more frequently referenced medical practices, single-surgeon websites, and government sources; *(iii)* according to Rothwell's classification system, technical details were the most frequently addressed topic on both platforms; and *(iv)* while evaluating the adequacy and accuracy of the responses, the majority were satisfactory; however, a substantial proportion still required moderate to significant clarification. Taken together, our results indicate that both ChatGPT-4o and Google offer significant academic answers, including a markedly high proportion of academic sources for those seeking information on RATHA.

Furthermore, we attempted to analyze the most FAQs about RATHA across two major online platforms, ChatGPT and Google, and to assess the informational quality and clinical relevance of the responses each platform offers. Unlike direct question-answer validation studies, this study focused on analyzing the thematic nature of questions and the patterns of content delivery by each platform.

A fundamental conceptual difference between the two platforms must be acknowledged. Google functions as a search engine that indexes and displays web content based on user queries. It does not provide answers per se but rather guides users to external content. ChatGPT, on the other hand, generates structured and cohesive textual responses derived from its language model training, presenting a more conversational and synthesized delivery of information.

The ChatGPT-4o is an AI-derived large language model (LLM) which generates realistic human responses via a chatbot function. It is trained via supervised and reinforcement learning to optimize the accuracy, breadth, and relevance of responses to text prompts using billions of modeling parameters and information obtained primarily from contemporary Internet sources.[14] Google Search Engine was selected as the control case, as it is the most widely used search engine worldwide and the only search engine that generates FAQs when prompted by a query. The FAQs were specifically selected for study, as *(i)* they are the most FAQs and, thus, of greatest interest to patients; *(ii)* this

allows for objective evaluation without bias from the authors in question generation; and *(iii)* this provides a systematic and reproducible method of question generation for comparison between Google Search Engine and ChatGPT-4o.

The prevalence of Internet use for health information among adult patients is a significant phenomenon in contemporary health behavior. Previous studies have indicated that over 60% of adults utilize the Internet to seek health information, reflecting its critical role as a resource for health-related inquiries.[15-17] This trend highlights a shift in how patients approach their health and the medical advice they receive, with many viewings' online resources as a viable supplement to traditional healthcare encounters. The evaluation of the quality of health information available on the Internet has become increasingly critical as more patients use online resources to make informed decisions regarding their health. Existing literature reveals considerable variability in the quality of health information across different websites.[18-21] Similar to studies in the literature assessing the quality of health information, this study compared ChatGPT's resources with Google Search engine's Robot-assisted THA FAQs. While the Internet can enhance communication and understanding, the quality of information varies, necessitating the involvement of healthcare providers to direct patients to credible resources and mitigate the risks associated with misinformation.

Our study revealed that 20% of the FAQs were similar between these two sources. Megalla et al.[22] reported that 30% of questions were similar between what Google and ChatGPT deemed to be the most FAQs. A prior research comparing Google and ChatGPT for total joint arthroplasty revealed that only 25% of the FAQs were similar across the two search engines.[11] In this respect, our findings are consistent with the existing literature. This limited overlap suggests that Google and ChatGPT offer distinct informational perspectives, which may complement each other in supporting patient education.

The ChatGPT provides a high percentage of academic resources as a reliable supplementary resource for patients seeking information from online sources. Dubin et al.[11] evaluated ChatGPT using Google FAQs and found ChatGPT to be a potential source of information for total hip and knee arthroplasty, and ChatGPT provided significantly more academic references than Google web search. Similarly, Tharakan et al.[23] compared Google and ChatGPT on total shoulder and elbow arthroplasty and found that both sources provided reliable information on these topics, but ChatGPT was the more reliable source from an academic and medical practice perspective. Varady et al.[24] found that ChatGPT-4 used a greater proportion of academic sources than Google to provide answers to the top 10 FAQs about ulnar collateral ligament. Moreover, another study revealed that ChatGPT-4 demonstrated the ability to provide accurate and reliable information about the Latarjet procedure in response to patient queries, using multiple academic sources in all cases and in contrast to Google Search Engine, which more frequently used single-surgeon and large medical practice websites.[25] In our study, ChatGPT provided significantly more academic references than Google search (70% *vs.* 20%). This finding is consistent with previous studies in the existing literature, where ChatGPT provided a high percentage of academic sources as a reliable additional resource for patients seeking information from online sources. Information from non-academic sources, such as commercial web pages and social media sites, might not be as accurate or unbiased as information from academic sources. Resources with a good reputation, such as academic journals and government websites, tend to be reliable and offer scientifically validated information.

There are several previous studies that have used the Rothwell classification to evaluate online queries and the quality of search engine results related to hip, knee, shoulder, and elbow arthroplasty. Dubin et al.[11] studied hip and knee arthroplasty and found that the most common subcategory was 'specific activities' (16 of 40), whereas 'technical details' were much less frequent (3 of 40). In their study on shoulder and elbow arthroplasty, Tharakan et al.[23] identified 'indications/management' as the most frequently addressed subcategory. Shen et al.[26] reported that the most popular question topics were 'Specific Activities' and 'Indications/Management'. In another study, McCormick et al.[27] also found 'Specific Activities' and 'Indications/Management' to be the most frequent Rothwell subcategories in a web-based analysis of FAQs related to arthroplasty. In contrast to the aforementioned studies in the literature, the most common subcategory by topic in our study was technical details. Given that our study focused on robot-assisted THA, this finding may be meaningful, as patients are likely more curious about the technical aspects of robotic surgery compared to conventional arthroplasty,

reflecting increased public interest in robotic technologies.

According to the results of our study, the absence of questions regarding pain, implant longevity, and surgical evaluation on both platforms may reflect user priorities or search behaviors at the time of query; however, it should not be interpreted as a definitive indicator of gaps in patient education.

The most concerning finding is that ChatGPT provided fabricated references for three of the questions. These fabricated references were shown as links that investigated similar to real references. They led users to wrong or nonexistent sources of information. On the other hand, Google never presented fabricated or incorrect links. Previous studies in the literature have also noted that ChatGPT shows fabricated references, tends to incorrectly suggest evidence, and fails to indicate when there is insufficient evidence to make a correct recommendation.[28,29] Therefore, both physicians and patients should be mindful that ChatGPT is experiencing "hallucinations" and should check these sources appropriately, as they are known to present false testimonials and information that may be false.[30]

Both conventional search engines and machine learning algorithms are expected to remain essential data sources of information for patients. However, to transform these data into meaningful insights, the adequacy and accuracy of the sources must be critically evaluated. Numerous studies in the literature have evaluated the adequacy and accuracy of responses provided by online sources such as Google and ChatGPT.[22,31-41] The outcomes reported in the existing literature demonstrate considerable variability. While some studies found the answers satisfactory,[22,32-35,37-39] others found them lacking.[31,36,40,41] In our study, similar to the methodologies employed in previous literature, we assessed the adequacy and accuracy of the responses of the FAQs. Responses classified as unsatisfactory were those that were inaccurate, outdated, or overly vague. In contrast, satisfactory responses were accurate but needed either minimal or moderate additional detail. Based on our analysis, 62% of FAQ responses were graded as 1 or 2, whereas 38% were classified as Grade 3 or 4. This underscores the fact that, despite the majority being satisfactory, a substantial proportion still necessitated moderate to significant clarification. Our findings align with the variability observed in previous studies. In orthopedic literature, robotic hip arthroplasty represents a relatively recent alternative to conventional techniques, which may explain our observation of information requiring significant clarification. Both Google and ChatGPT-4o need further refinement to ensure the reliability of information in the field of robotic hip arthroplasty. Given the rapid evolution of these models, continuous reassessment is essential. Developing new and comprehensive tools to evaluate the quality and accuracy of medical information is crucial to enable these models to effectively support patient education. Future research should focus on improving the adequacy and accuracy of information to better serve patients.

Educating patients on how to get information from Internet sources is crucial to reduce misunderstanding and misinformation.[42] Healthcare professionals should recognize that ChatGPT and Google probably use identical sources for a specific inquiry.[43] The key difference is that ChatGPT synthesizes information from multiple sources to arrive at a single answer, while Google maintains its uniqueness by presenting a multitude of results. In subjects characterized by low consensus and, therefore, a lack of reliable sources, there exists a significantly elevated likelihood that ChatGPT would reference less accurate material. In such instances, physicians have to invest time in educating patients on the subject or supplying resources that provide more reliable information. The main goal of AI is not to replace healthcare professionals, but to improve the patient experience and help physicians make better decisions, which would make patients safer and make physicians more reliable and efficient.[44]

Nonetheless, this study has several limitations that should be acknowledged. First, although both platforms yielded 20 FAQs, these were not identical, which prevents direct, question-by-question comparison. Second, the small sample size (n=40 total questions) may limit the generalizability of our findings. Patients may use a broader range of search terms beyond those tested in this study, potentially resulting in different queries and outputs. Additionally, we did not assess readability or patient-oriented clinical usefulness of the responses, which represent important dimensions of information quality. Future research should incorporate these parameters to provide a more comprehensive evaluation.

Google's dynamic and personalized search algorithms, affected by user history, location, and device, may have introduced variability into the search results, despite efforts to minimize this

using a clean browser. Furthermore, while we analyzed the thematic content and source types of responses, we did not formally assess the medical accuracy, depth, or comprehensibility of the answers, which are essential dimensions for evaluating the platforms' effectiveness in patient education.

Finally, although ChatGPT-4o was used to generate responses, it is of utmost importance to note that its knowledge base is not updated in real time. As of this study, its data only reflects information available until January 2025. This temporal lag may limit its ability to provide up-to-date clinical recommendations.

For future studies expanding the research to cover a broader range of questions and evaluating the quality of responses would offer more in-depth insights. Exploring how patients combine information from AI and conventional sources could help shape the creation of integrated patient education strategies.

In conclusion, ChatGPT serves as a valuable alternative to traditional search engines for patients seeking information about RATHA. Our study results revealed that ChatGPT provided more academic references than Google. According to the Rothwell classification, technical details were found to be the most frequent subcategory, indicating interest in the content of robot technology. While assessing the accuracy of the responses, we observed a considerable proportion of information that required moderate to significant clarification. Given the increasing reliance on online platforms for medical information, ChatGPT may serve as a clinical adjunct under the supervision of a physician when addressing questions on RATHA.

Data Sharing Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions: .Writing, editing, analysis: M.F.D.; Writing, data collection, statistics: S.S.; Data collection, analysis: F.A.; Data collection, proofreading, editing: M.A.; Analysis, statistics, writing, editing: C.T.; Data collection, analysis, statistics: N.A.S.; Supervision, editing, proofreading: M.C. All authors contributed to the study conception and design. 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. Yin H, Zhang Y, Hou W, Wang L, Fu X, Liu J. Comparison of complications between total hip arthroplasty following failed internal fixation and primary total hip arthroplasty for femoral neck fractures: A meta-analysis. Jt Dis Relat Surg 2025;36:479-88. doi: 10.52312/jdrs.2025.2230.
2. Tsertsvadze A, Grove A, Freeman K, Court R, Johnson S, Connock M, et al. Total hip replacement for the treatment of end stage arthritis of the hip: A systematic review and meta-analysis. PLoS One 2014;9:e99804. doi: 10.1371/journal.pone.0099804.
3. Wang Y, Wang R, Gong S, Han L, Gong T, Yi Y, et al. A comparison of radiological and clinical outcomes between robotic-assisted and conventional total hip arthroplasty: A meta-analysis. Int J Med Robot 2023;19:e2463. doi: 10.1002/rcs.2463.
4. Nawabi DH, Conditt MA, Ranawat AS, Dunbar NJ, Jones J, Banks S, et al. Haptically guided robotic technology in total hip arthroplasty: A cadaveric investigation. Proc Inst Mech Eng H 2013;227:302-9. doi: 10.1177/0954411912468540.
5. Kim K, Kwon S, Kwon J, Hwang J. A review of robotic-assisted total hip arthroplasty. Biomed Eng Lett 2023;13:523-35. doi: 10.1007/s13534-023-00312-9.
6. Atik OŞ. Does the use of robotic technology in hip arthroplasty provide superior clinical outcomes? Jt Dis Relat Surg 2022;33:253-4. doi: 10.52312/jdrs.2022.57904.
7. 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.
8. Pan A, Musheyev D, Bockelman D, Loeb S, Kabarriti AE. Assessment of artificial intelligence chatbot responses to top searched queries about cancer. JAMA Oncol 2023;9:1437-40. doi: 10.1001/jamaoncol.2023.2947.
9. Chakraborty C, Pal S, Bhattacharya M, Dash S, Lee SS. Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Front Artif Intell 2023;6:1237704. doi: 10.3389/frai.2023.1237704.
10. Andrikyan W, Sametinger SM, Kosfeld F, Jung-Poppe L, Fromm MF, Maas R, et al. Artificial intelligence-powered chatbots in search engines: A cross-sectional study on the quality and risks of drug information for patients. BMJ Qual Saf 2025;34:100-9. doi: 10.1136/bmjqs-2024-017476.
11. Dubin JA, Bains SS, Chen Z, Hameed D, Nace J, Mont MA, et al. Using a Google Web search analysis to assess the utility of ChatGPT in total joint arthroplasty. J Arthroplasty 2023;38:1195-202. doi: 10.1016/j.arth.2023.04.007.
12. Rothwell JD. In mixed company: Communicating in small groups. Boston, MA: Cengage Learning. Account Educ 2013. p. 23.
13. Mika AP, Martin JR, Engstrom SM, Polkowski GG, Wilson JM. Assessing ChatGPT responses to common patient questions regarding total hip arthroplasty. J Bone Joint Surg Am 2023;105:1519-26. doi: 10.2106/JBJS.23.00209.
14. Kunze KN, Jang SJ, Fullerton MA, Vigdorchik JM, Haddad FS. What's all the chatter about? Bone Joint J 2023;105-B:587-9. doi: 10.1302/0301-620X.105B6.BJJ-2023-0156.
15. Jabson JM, Patterson JG, Kamen C. Understanding health information seeking on the internet among sexual minority people: Cross-sectional analysis from the Health Information National Trends Survey. JMIR Public Health Surveill 2017;3:e39. doi: 10.2196/publichealth.7526.

16. Giorgino R, Alessandri-Bonetti M, Del Re M, Verdoni F, Peretti GM, Mangiavini L. Google Bard and ChatGPT in orthopedics: Which is the better doctor in sports medicine and pediatric orthopedics? The role of AI in patient education. Diagnostics (Basel) 2024;14:1253. doi: 10.3390/diagnostics14121253.
17. Estanislau AR, Nobre DM, Naback FDN, Barros TSV, Souza LFM, Penido PCL. Use of the internet and social networks in orthopedics and traumatology and perspective of post COVID telemedicine. Acta Ortop Bras 2022;30:e252728. doi: 10.1590/1413-785220223005e252728.
18. Kaicker J, Borg Debono V, Dang W, Buckley N, Thabane L. Assessment of the quality and variability of health information on chronic pain websites using the DISCERN instrument. BMC Med 2010;8:59. doi: 10.1186/1741-7015-8-59.
19. Dobbins M, Watson S, Read K, Graham K, Yousefi Nooraie R, Levinson AJ. A tool that assesses the evidence, transparency, and usability of online health information: Development and reliability assessment. JMIR Aging 2018;1:e3. doi: 10.2196/aging.9216.
20. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: Systematic literature review of relevant criteria and indicators. J Med Internet Res 2019;21:e12522. doi: 10.2196/12522.
21. Vivekanantham A, Protheroe J, Muller S, Hider S. Evaluating on-line health information for patients with polymyalgia rheumatica: A descriptive study. BMC Musculoskelet Disord 2017;18:43. doi: 10.1186/s12891-017-1416-5.
22. Megalla M, Hahn AK, Bauer JA, Windsor JT, Grace ZT, Gedman MA, et al. ChatGPT and Google provide mostly excellent or satisfactory responses to the most frequently asked patient questions related to Rotator Cuff Repair. Arthrosc Sports Med Rehabil 2024;6:100963. doi: 10.1016/j.asmr.2024.100963.
23. Tharakan S, Klein B, Bartlett L, Atlas A, Parada SA, Cohn RM. Do ChatGPT and Google differ in answers to commonly asked patient questions regarding total shoulder and total elbow arthroplasty? J Shoulder Elbow Surg 2024;33:e429-37. doi: 10.1016/j.jse.2023.11.014.
24. Varady NH, Lu AZ, Mazzucco M, Dines JS, Altchek DW, Williams RJ 3rd, et al. Understanding how ChatGPT may become a clinical administrative tool through an investigation on the ability to answer common patient questions concerning Ulnar Collateral Ligament Injuries. Orthop J Sports Med 2024;12:23259671241257516. doi: 10.1177/23259671241257516.
25. Oeding JF, Lu AZ, Mazzucco M, Fu MC, Taylor SA, Dines DM, et al. ChatGPT-4 performs clinical information retrieval tasks using consistently more trustworthy resources than does Google Search for queries concerning the latarjet procedure. Arthroscopy 2025;41:588-97. doi: 10.1016/j.arthro.2024.05.025.
26. Shen TS, Driscoll DA, Islam W, Bovonratwet P, Haas SB, Su EP. Modern internet search analytics and total joint arthroplasty: What are patients asking and reading online? J Arthroplasty 2021;36:1224-31. doi: 10.1016/j.arth.2020.10.024.
27. McCormick JR, Kruchten MC, Mehta N, Damodar D, Horner NS, Carey KD, et al. Internet search analytics for shoulder arthroplasty: What questions are patients asking? Clin Shoulder Elb 2023;26:55-63. doi: 10.5397/cise.2022.01347.
28. Subramanian T, Shahi P, Araghi K, Mayaan O, Amen TB, Iyer S, et al. Using artificial intelligence to answer common patient-focused questions in minimally invasive spine surgery. J Bone Joint Surg Am 2023;105:1649-53. doi: 10.2106/JBJS.23.00043.
29. Shrestha N, Shen Z, Zaidat B, Duey AH, Tang JE, Ahmed W, et al. Performance of ChatGPT on NASS Clinical Guidelines for the diagnosis and treatment of low back pain: A comparison study. Spine (Phila Pa 1976) 2024;49:640-51. doi: 10.1097/BRS.0000000000004915.
30. Price WN 2nd, Gerke S, Cohen IG. How much can potential jurors tell us about liability for medical artificial intelligence? J Nucl Med 2021;62:15-6. doi: 10.2967/jnumed.120.257196.
31. Johnson CK, Mandalia K, Corban J, Beall KE, Shah SS. Adequacy of ChatGPT responses to frequently asked questions about shoulder arthroplasty: Is it an appropriate adjunct for patient education? JSES Int 2025;9:830-6. doi: 10.1016/j.jseint.2025.01.008.
32. 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 2025;41:1806-27. doi: 10.1016/j.arthro.2024.08.036.
33. Slawaska-Eng D, Bourgeault-Gagnon Y, Cohen D, Pauyo T, Belzile EL, Ayeni OR. ChatGPT-3.5 and -4 provide mostly accurate information when answering patients' questions relating to femoroacetabular impingement syndrome and arthroscopic hip surgery. J ISAKOS 2025;10:100376. doi: 10.1016/j.jisako.2024.100376.
34. Adelstein JM, Sinkler MA, Li LT, Fortier LM, Vakharia AM, Salata MJ. ChatGPT can often respond adequately to common patient questions regarding femoroacetabular impingement. Clin J Sport Med 2024. doi: 10.1097/JSM.0000000000001327.
35. Adelstein JM, Sinkler MA, Li LT, Mistovich RJ. ChatGPT responses to common questions about slipped capital femoral epiphysis: A reliable resource for parents? J Pediatr Orthop 2024;44:353-7. doi: 10.1097/BPO.0000000000002681.
36. Li AW, Adelstein JM, Li LT, Sinkler MA, Mistovich RJ. Assessing ChatGPT responses to frequently asked questions regarding pediatric supracondylar humerus fractures. J Pediatr Orthop 2025;45:327-31. doi: 10.1097/BPO.0000000000002923.
37. AlShehri Y, McConkey M, Lodhia P. ChatGPT provides satisfactory but occasionally inaccurate answers to common patient hip arthroscopy questions. Arthroscopy 2025;41:1337-47. doi: 10.1016/j.arthro.2024.06.017.
38. Shayegh NA, Byer D, Griffiths Y, Coleman PW, Deane LA, Tonkin J. Assessing artificial intelligence responses to common patient questions regarding inflatable penile prostheses using a publicly available natural language processing tool (ChatGPT). Can J Urol 2024;31:11880-5.
39. Winden F, Bormann M, Gilbert F, Holzapfel BM, Berthold DP. ChatGPT delivers satisfactory responses to the most frequent questions on meniscus surgery. Knee 2025;56:249-57. doi: 10.1016/j.knee.2025.05.018.
40. Li LT, Adelstein JM, Sinkler MA, Mistovich RJ. Artificial intelligence promotes the Dunning Kruger Effect: Evaluating ChatGPT answers to frequently asked questions about adolescent idiopathic scoliosis. J Am Acad Orthop Surg 2025;33:473-80. doi: 10.5435/JAAOS-D-24-00297.

41. Kolac UC, Karademir OM, Ayik G, Kaymakoglu M, Familiari F, Huri G. Can popular AI large language models provide reliable answers to frequently asked questions about rotator cuff tears? JSES Int 2024;9:390-7. doi: 10.1016/j.jseint.2024.11.012.
42. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: A systematic review. J Med Internet Res 2017;19:e9. doi: 10.2196/jmir.5729.
43. Shen OY, Pratap JS, Li X, Chen NC, Bhashyam AR. How does ChatGPT use source information compared with Google? A text network analysis of online health information. Clin Orthop Relat Res 2024;482:578-88. doi: 10.1097/CORR.0000000000002995.
44. Atik OŞ, Sezgin EA, Bahadır B. The main goal of ChatGPT is not to replace healthcare professionals. Jt Dis Relat Surg 2024;35:471-2. doi: 10.52312/jdrs.2024.57923.