

## Research Article

# Artificial Intelligence Provides Helpful Information to Patients Undergoing ACLR, But Patients May Not Fully Understand The Answers

Grace Basralian, BS<sup>1</sup><sup>a</sup>, Kevin M. Posner, BS, MD<sup>1b</sup>, Grace Chester, BS<sup>1</sup>, Mallery Zeiman, BS<sup>1c</sup>, Yair D. Kissin, MD<sup>2d</sup>, Stephen G. Silver, MD<sup>2e</sup>, Amit Y. Merchant, DO<sup>2f</sup>

<sup>1</sup> Orthopedics, Hackensack Meridian Health, <sup>2</sup> Orthopedics, Hackensack University Medical Center

Keywords: Artificial intelligence, ACL reconstruction, Patient education, Readability

<https://doi.org/10.60118/001c.145736>

---

## Journal of Orthopaedic Experience & Innovation

Vol. 7, Issue 1, 2026

---

a Grace Basralian is a medical student at Hackensack Meridian School of Medicine (Class of 2026), where she was awarded the Dean's Award for Research and Scholarship. She earned her Bachelor of Science in Exercise Science from George Washington University, graduating summa cum laude. Her research focuses on orthopedic surgery and related fields, with multiple publications in peer-reviewed journals including Cureus, Urology Case Reports, SAGE Digital Health, and the Archives of Bone and Joint Surgery. She has presented her work nationally at conferences such as Shoulder360, the American Urological Association, and the Society for Maternal-Fetal Medicine. Her scholarly interests include orthopedics, artificial intelligence in medicine, and clinical outcomes research.

[Connect with Grace Basralian on LinkedIn](#)

[Conflicts of Interest Statement for Grace Basralian](#)

b Dr. Kevin Posner is an Orthopedic Surgery Resident at NYU Langone Health, passionate about orthopedic surgery and committed to advancing patient care through evidence-based practice. He earned his MD from Hackensack Meridian School of Medicine and his BS in Exercise Physiology from the University of Miami.

[Conflicts of Interest Statement for Dr. Kevin Posner](#)

c [Conflicts of Interest Statement for Mallery Zeiman](#)

d Dr. Kissin is a knee specialist, with a particular focus on anterior cruciate ligament tears and restoration, and robotic total knee replacement. He is Vice chair of the department at Hackensack University, Medical Center and involved in the residency program. He is also associate professor at Hackensack school of medicine and Rutgers Newark.

[Visit Dr. Kissin](#)

[Connect with Dr. Kissin on LinkedIn](#)

[Conflicts of Interest Statement for Dr. Kissin](#)

[Visit the Open Payments Data Page for Dr. Kissin](#)

e Stephen Silver, M.D., Chief, Department of Sports Medicine at Hackensack University Medical Center, is a board-certified orthopedic surgeon specializing in arthroscopic treatment of the shoulder, knee, and elbow. He has extensive experience in sports medicine and is thoughtful, meticulous and thorough in his approach. He makes it a priority to foster comfort and confidence in his patients by helping them understand their diagnosis and the range of available treatment options. His goal is to not only restore, but improve quality of life for all his patients. Dr. Silver has been serving the community of northern New Jersey for over 20 years. He has spent 10 seasons as NBA team physicians for the New Jersey Nets, now known as the Brooklyn Nets. Dr. Silver has been responsible for a number of innovative advances in rotator cuff repair and posterior cruciate ligament (PCL) reconstruction, and has been recognized as an Orthopedic Top Doctor by Castle Connolly and by 201 Magazine and Bergen Magazine for the past 10 years.

[Visit Dr. Stephen Silver](#)

[Connect with Dr. Stephen Silver on LinkedIn](#)

[Conflicts of Interest Statement for Dr. Stephen Silver](#)

[Visit the Open Payments Data Page for Dr. Stephen Silver](#)

f Amit Merchant, D.O., is board certified in orthopedic surgery and specializes in pediatric orthopaedics, caring for children from birth to adolescence. His areas of clinical interest include scoliosis, hip dysplasia, angular deformities, neuromuscular disorders, pediatric trauma, sports related injuries.

[Visit Dr. Amit Merchant](#)

[Conflicts of Interest Statement for Dr. Amit Merchant](#)

[Visit the Open Payments Data Page for Dr. Amit Merchant](#)

## Background

With the increasing integration of medicine and technology, evaluating the accuracy of online resources and newly introduced artificial intelligence (AI) platforms is critical. Despite the numerous chatbot systems available, there is little information regarding their reliability in providing accurate information about common orthopaedic procedures, such as anterior cruciate ligament reconstruction (ACLR).

**Purpose/Hypothesis:** This study aims to compare the accuracy and readability of information provided by OpenAI's ChatGPT and Google's Gemini in response to commonly asked patient questions prior to ACLR. The authors hypothesize that AI systems will be able to provide useful information, but the information will not be suited for patient

## Methods

Authors compiled a list of common questions directed towards orthopaedic surgeons prior to ACLR and inputted the questions into ChatGPT 3.5 and Google's Gemini. Using a four-point scale ranging from one being an excellent response to four being an unsatisfactory response needing major clarification, a group of board-certified orthopaedic surgeons independently graded each system's answers. Each surgeon then provided an overall answer as to whether they felt the chat system's response would be helpful to patients. Additionally, each response was evaluated for its readability using the Flesch-Kincaid Reading Ease and Flesch-Kincaid Grade Level.

## Results

Both ChatGPT and Gemini provided responses to all 13 questions, each including disclaimers recommending consultation with medical professionals. Surgeon grading demonstrated that ChatGPT responses received average scores ranging from 1.15 to 2.23, while Gemini responses ranged from 1.31 to 3.15. Significant variability was noted in the scores provided to ChatGPT responses across each question ( $F = 4.13$ ,  $p = 0.00016$ ) and among surgeons ( $F = 3.18$ ,  $p = 0.0196$ ), whereas Gemini scores were more consistent across the responses to each question but again showed variability between surgeons ( $F = 13.11$ ,  $p < 0.001$ ). Readability analysis favored Gemini, with a Flesch-Kincaid Ease score of 43.27 compared to 23.5 with ChatGPT ( $p < 0.001$ ). The average grade level required to understand the response provided by ChatGPT was 12, while the average grade level required to understand Gemini's response was 11.06 ( $p < 0.001$ ).

## Conclusion

ChatGPT and Gemini were both able to provide adequate, helpful responses to patient questions with respect to ACLR, as graded by board-certified orthopaedic surgeons. Despite this the responses required a reading comprehension level well above the average level for an American which is 7th/8th grade level, raising concerns regarding patient comprehension. It is essential for the medical community to continue critically evaluating AI-generated information to ensure it is accurate, reliable, and easily understandable for patients.

## INTRODUCTION

When facing a surgical procedure, it is common for patients to have questions regarding details of their procedure, recovery and/or the rehabilitation process. While some patients may already possess background knowledge on their surgery, many will not and may be undergoing a procedure for the first time in their life. Even worse, many patients commonly refer to misleading sources about their diagnosis and treatment options, which can lead to confusion and obligate the care giver to explain fact from fiction. When examining where patients seek out health information, much of the time their searching is centered around commercial websites or search engines as opposed to academic or scholarly journals (LaValley, Kiviniemi, and Gage-Bouchard

2017). With the surge of information regarding orthopedic injuries on social media, patients may even resort to platforms such as Facebook, X (formerly Twitter), Instagram, or Tik Tok as a means to answer unmet clinical questions that may arise (Hasley, Bukowiec, Zaifman, et al. 2023). With the recent surge of artificial intelligence (AI) into mainstream use, it begs the question, how can this new technology, such as the widely available platform ChatGPT or Google Gemini, be utilized in medicine?

ChatGPT in particular is an AI platform that is constructed so that the user can propose questions to the system and receive a response in a conversational manner. ChatGPT formulates its response by analyzing all public, free information available on the internet including but not limited to, books, websites, and articles and forming an ar-

tificial neuron network to provide a reply in a humanistic manner (Cascella et al. 2023). The system is able to go back and forth with the user, answering follow up questions and clarifying information while fostering an on-going dialogue. Similarly, Google's Gemini, another prominent and free AI platform, offers comparable capabilities leveraging its extensive access to information to provide real-time answers to user queries. With both platforms available for public use, it raises the question of how these technologies might assist in answering common medical questions, particularly in the context of surgery and rehabilitation.

In the field of medicine ChatGPT has been proven effective at answering general medical queries (Johnson, Goodman, Patrinely, et al. 2023). In addition to this ChatGPT demonstrated that it was capable of performing at or near the threshold needed to pass a National Board of Medical Examiners (NBME) style exam and was equated to that of a third year medical student (Kung, Cheatham, Medenilla, et al. 2023; Gilson, Safranek, Huang, et al. 2023). Furthermore, researchers demonstrated that AI chatbot responses may even be superior to those of a licensed medical provider. One study demonstrated that the responses from an AI chatbot to general medical questions were actually preferred over those of a verified medical provider when being evaluated by other medical professionals (Ayers, Poliak, Dredze, et al. 2023). Studies evaluating Google's Gemini have shown comparable results to ChatGPT in terms of accuracy when answering medical questions, including those related to pediatric orthopaedics (Pirkle, Yang, and Blumberg 2024). However, occasional discrepancies were noted, particularly in Gemini's citations, where several referenced studies were found to contain errors or were unverifiable (Pirkle, Yang, and Blumberg 2024).

AI has exceeded what was previously thought possible with respect to internet chat systems. The success of AI systems and their easy accessibility raise the question, are patients using AI to answer their medical questions? For orthopaedics, there are a number of questions that a patient may have, particularly in regard to potential surgical procedures. Anterior cruciate ligament reconstruction (ACLR) is a common orthopaedic procedure that is often performed on younger patients (Leathers et al. 2015). This information, coupled with the knowledge that younger patients often turn to online resources for health related information, creates the need to evaluate online health resources (Jia, Pang, and Liu 2021). The aim of this study is twofold: (1) to evaluate whether ChatGPT and Google's Gemini can be relied upon as a resource for patients to access accurate information regarding preoperative, intraoperative, and post-operative events in relation to common orthopaedic procedures, particularly ACLR; and (2) to compare the accuracy of information provided against each other. Previous studies have analyzed the usability and accuracy of ChatGPT for various orthopedic conditions. To our knowledge no prior studies have comparatively evaluated AI platforms against each other for accuracy and comprehension regarding ACLR.

## METHODS

Authors formulated thirteen potential patient questions that may arise prior to ACLR. Each question was then reviewed by a group of five board-certified orthopedic surgeons and their validity was confirmed for inclusion. Questions were then entered into both ChatGPT 3.5 and Google's Gemini (Table 1). Responses were recorded and de-identified so that the AI system of origin was not discernible. The questions, along with the de-identified answers, were then independently evaluated by the same group of board-certified orthopedic surgeons. A four-point scale, with a score of one representing an excellent response and four indicating a need for major clarification, was used to grade each AI system's response (Li et al. 2024; Johns et al. 2024). Additionally, the surgeons indicated with a 'yes' or 'no' whether they believed each AI system's response would be helpful to patients. For example, one response to the question "Do I need ACL reconstruction surgery?," ChatGPT responded with "However, whether or not you need ACL (anterior cruciate ligament) reconstruction surgery depends on various factors including the severity of your injury, your activity level, your age, and your overall health. If you have experienced an ACL tear, it's essential to consult with a qualified orthopedic surgeon or sports medicine specialist. They will assess your condition through physical examination, imaging tests (such as MRI), and discussion of your symptoms and lifestyle. Based on their evaluation, they can recommend the most appropriate treatment plan, which may or may not involve surgery. In some cases, particularly if you're not highly active or if the tear is partial and your knee remains stable, conservative treatment such as physical therapy and bracing may be sufficient. However, if you're an athlete or if the tear is severe and causing instability in your knee, surgery may be recommended to reconstruct the torn ACL and restore stability to the joint.

Ultimately, the decision to undergo ACL reconstruction surgery should be made in consultation with a qualified medical professional who can provide personalized advice based on your individual circumstances." which all reviews marked as "yes." This indicated that the surgeons agreed with this response and they thought the response would be overall helpful for the patient to read. Table 2 shows examples of responses from both ChatGPT and Gemini showing the difference in how each AI answers the same question. For the first question all but one surgeon found ChatGPT's answer useful while all surgeons found Gemini's answer useful. However, for the second question, three of the surgeons found ChatGPT's answer not useful while all surgeons found Gemini's answer useful.

Each response was also evaluated for readability. The reading level of the ChatGPT and Google Gemini responses were calculated using the WordCalc software package, and readability was assessed using the Flesch-Kincaid (FK) grade level, FK reading ease score, Simple Measure of Gobbledygook index (SMOG), Coleman-Liau index (CL), Gunning fog index (GF), and automated readability index.

**Table 1. Proposed Questions**

Number	Question
1	Do I need ACL reconstruction surgery?
2	How long does ACL reconstruction surgery take?
3	What are the risks of ACL reconstruction surgery?
4	How long do I need a brace after ACL reconstruction surgery?
5	When can I walk after ACL reconstruction surgery?
6	Will I have pain after ACL reconstruction surgery?
7	How long until I can run after ACL reconstruction surgery?
8	How long until I can return to sports after ACL reconstruction surgery?
9	Why did I tear my ACL?
10	What if I choose not to have my ACL injury fixed?
11	Can my ACL be repaired?
12	Will I have arthritis because I tore my ACL?
13	Should I have an autograft or allograft used during ACL surgery?

## RESULTS

Both systems provided responses to all 13 (100.00%) and utilized what was determined to be logical reasoning without providing information that contraindicated itself. Additionally, 100% of the responses provided by both systems included a disclaimer to discuss further medical information with either a “medical team”, “physical therapist”, and/or “surgeon”.

When comparing the average scores for each question, ChatGPT had a lower average score, indicating a better response, on 9 of the 13 questions analyzed. ChatGPT’s lowest average score was on question 6 at 1.0, while Gemini’s lowest average score was for question 5 at 1.4. ChatGPT’s highest average score appeared on question 5 at 3.2, whereas Gemini’s highest average score was for question 1 at 2.8. Significant differences between the average score on questions 1 and 5 were noted ( $p < 0.01$ ), while other questions demonstrated comparable performance between ChatGPT and Gemini. Across all thirteen questions, three of the surgeons scored ChatGPT with lower scores on average compared to Gemini (Table 3). When comparing the average scores provided to ChatGPT and Gemini responses provided by each surgeon across the entire question set, surgeon 5 was noted to score ChatGPT responses significantly better (2.23 vs 3.15;  $p = 0.021$ ) (Table 3).

Overall, ChatGPT responses were deemed useful an average of 12 times by each surgeon, while Gemini responses were rated useful 11.4 times on average (Table 3). Surgeon 3 rated the highest number of ChatGPT responses as useful, with 13 out of 13 rated “yes,” while Surgeon 1 gave the fewest “yes” ratings for ChatGPT at 11 (Table 3). For Gemini, Surgeon 3 also rated all responses as useful (13 out of 13), while Surgeon 5 found the fewest Gemini responses useful, with only 9 “yes” ratings (Table 3).

## CHATGPT

ChatGPT responses demonstrated variability across questions and surgeons. With respect to each surgeon’s score on the set of responses the average ranged from 1.15 to 2.23, with surgeon one scoring an average of 1.85 (SD = 1.07), surgeon two 1.77 (SD = 0.83), surgeon three 1.15 (SD = 0.38), surgeon four 1.69 (SD = 0.63), and surgeon five scoring 2.23 (SD = 0.83). ANOVA analysis revealed significant variability in scores for responses to the same question ( $F = 4.13$ ,  $p = 0.00016$ ). Significant variability was also found in scores assigned by different surgeons across all responses ( $F = 3.18$ ,  $p = 0.0196$ ).

## GEMINI

When analyzing the scores assigned to each response by each surgeon, the average ranged from 1.31 to 3.15. Surgeon one averaged 1.62 (SD = 0.87), Surgeon two 1.77 (SD = 0.73), Surgeon three 1.31 (SD = 0.48), Surgeon four 2.00 (SD = 0.71), and Surgeon five 3.15 (SD = 0.69). ANOVA revealed no significant variability in scores for responses to the same question ( $F = 1.28$ ,  $p = 0.26$ ). Significant variability was found in scores assigned by different surgeons across all responses ( $F = 13.11$ ,  $p < 0.001$ ).

## READABILITY

ChatGPT’s Flesch-Kincaid (FK) Reading Ease score was 23.5, while Gemini scored 43.27 ( $p < 0.001$ ). ChatGPT’s New Dale Chall (NDC) Score was 8.19, compared to Gemini’s 7.27. Both systems recorded an Spache Readability (SR) Score of 5. FK Grade Level scores were 12 for ChatGPT and 11.06 for Gemini ( $p < 0.001$ ). Gunning Fog (GF) Index scores were 18.65 for ChatGPT and 14.92 for Gemini, with both systems recording Coleman-Liau (CL) scores of 12. SMOG index values were 0.83 for ChatGPT and 0.63 for Gemini.

**Table 2. Examples of Answer from both AI Systems**

Question	ChatGPT	Gemini
<p>How long until I can run after ACL reconstruction surgery?</p>	<p>The timeline for returning to running after ACL (anterior cruciate ligament) reconstruction surgery varies from person to person and depends on factors such as the extent of the surgery, individual healing capabilities, adherence to rehabilitation protocols, and guidance from your surgeon and physical therapist. Generally, here's what you can expect:</p> <p><b>Early Post-Operative Phase (First Few Weeks):</b> Running is typically not permitted during the initial phase of ACL rehabilitation, which focuses on protecting the surgical site, reducing swelling, regaining range of motion, and strengthening the muscles around the knee. During this period, you'll likely engage in gentle exercises, such as walking, stationary cycling, and low-impact activities, under the guidance of your physical therapist.</p> <p><b>Intermediate Phase (Weeks 6-12):</b> Around 6 to 12 weeks after surgery, you may gradually progress to more advanced rehabilitation exercises, including light jogging or running on a treadmill. However, the intensity and duration of running should be carefully controlled to avoid excessive stress on the healing tissues. Your physical therapist will design a structured program to gradually increase running intensity and volume while monitoring your progress and ensuring proper form.</p> <p><b>Advanced Rehabilitation Phase (Months 3-6 and Beyond):</b> By 3 to 6 months post-surgery, many patients are able to resume running at a higher intensity and duration, although this timeline can vary depending on individual factors. Your physical therapist and surgeon will assess your readiness to return to running based on factors such as strength, stability, range of motion, and functional performance.</p> <p><b>Full Return to Running:</b> A full return to running, including activities such as sprinting, cutting, and agility drills, typically occurs around 6 to 9 months after ACL reconstruction surgery. However, it's essential to progress gradually and listen to your body to avoid overloading the knee or risking re-injury.</p> <p>It's crucial to follow your surgeon's and physical therapist's recommendations closely regarding the timing and progression of running activities after ACL reconstruction surgery. Returning to running too soon or without proper guidance can increase the risk of complications, re-injury, or long-term issues. Prioritize a gradual and structured approach to rehabilitation to optimize your recovery and minimize the risk of setbacks.</p>	<p>You can actually start walking fairly soon after ACL reconstruction surgery, though with some limitations:</p> <p><b>1-2 days post-surgery:</b> Most patients are able to walk short distances with the help of crutches [various sources]. This walking is usually for strengthening and reducing swelling, not for getting around independently.</p> <p><b>2-4 weeks post-surgery:</b> The goal of physical therapy during this time is to gradually increase weight-bearing and wean off crutches as your strength improves. By this stage, you should be able to walk for short distances without crutches [various sources].</p> <p>It's important to remember these are general guidelines. The specific timeframe for when you can walk unassisted will depend on several factors, including:</p> <p><b>The severity of your injury:</b> More complex tears or additional ligament damage might require a longer period using crutches for support.</p> <p><b>Your physical therapy progress:</b> Your dedication to physical therapy exercises significantly impacts how quickly you regain strength and stability in your knee. Your surgeon's instructions: Follow your doctor's specific recommendations regarding weight-bearing and walking after surgery.</p> <p>Here are some additional things to keep in mind:</p> <p>Walking will likely be painful at first. Pain management medication can help, and physical therapy will focus on improving your gait and reducing discomfort.</p> <p>Use of a knee brace might be recommended initially. This can provide support and stability as you walk.</p> <p>Walking will gradually increase in distance and duration as you progress through physical therapy. It's always best to discuss your recovery timeline, including when you can expect to walk without limitations, with your doctor or physical therapist. They can provide a personalized plan based on your individual circumstances.</p>
<p>When can I walk after ACL reconstruction surgery?</p>	<p>The timeline for walking after ACL (anterior cruciate ligament) reconstruction surgery varies depending on several factors, including the specific surgical technique used, individual healing capabilities, and the guidance of your surgeon and physical therapist. However, here's a general overview of what to expect:</p> <p><b>Immediately After Surgery:</b> You will likely not be able to bear weight on the affected leg immediately after ACL reconstruction surgery. You may require crutches or a walker to assist with mobility.</p> <p><b>First Few Days:</b> In the days immediately following surgery, you'll need to keep weight off the operated leg to allow the initial healing</p>	<p>You can actually start walking fairly soon after ACL reconstruction surgery, though with some limitations:</p> <p><b>1-2 days post-surgery:</b> Most patients are able to walk short distances with the help of crutches [various sources]. This walking is usually for strengthening and reducing swelling, not for getting around independently.</p> <p><b>2-4 weeks post-surgery:</b> The goal of</p>

Question	ChatGPT	Gemini
	<p>process to occur. During this time, you may use crutches or a walker for support when moving around.</p> <p>First Few Weeks: Your surgeon will provide specific instructions regarding when you can start bearing weight on the operated leg. Typically, weight-bearing activities such as walking with partial or full weight on the leg may begin within the first few weeks after surgery, depending on the extent of the procedure and your individual progress.</p> <p>Gradual Progression: As you continue to heal and regain strength in the operated leg, you'll gradually increase weight-bearing activities and walking without assistance. Your physical therapist will guide you through a structured rehabilitation program to ensure a safe and effective recovery.</p> <p>Rehabilitation Phase: Walking is an essential component of ACL rehabilitation, and it will be incorporated into your therapy program as you progress. Initially, you may walk with a limp or experience some discomfort, but over time, your gait should improve as your knee heals and your strength and stability increase.</p> <p>It's crucial to follow your surgeon's and physical therapist's recommendations closely regarding weight-bearing and walking activities after ACL reconstruction surgery. Pushing yourself too hard or returning to activities too quickly can increase the risk of complications or re-injury. It's essential to prioritize proper healing and rehabilitation to optimize the long-term outcome of your surgery."</p>	<p>physical therapy during this time is to gradually increase weight-bearing and wear off crutches as your strength improves. By this stage, you should be able to walk for short distances without crutches [various sources].</p> <p>It's important to remember these are general guidelines. The specific timeframe for when you can walk unassisted will depend on several factors, including:</p> <p>The severity of your injury: More complex tears or additional ligament damage might require a longer period using crutches for support.</p> <p>Your physical therapy progress: Your dedication to physical therapy exercises significantly impacts how quickly you regain strength and stability in your knee. Your surgeon's instructions: Follow your doctor's specific recommendations regarding weight-bearing and walking after surgery.</p> <p>Here are some additional things to keep in mind:</p> <p>Walking will likely be painful at first. Pain management medication can help, and physical therapy will focus on improving your gait and reducing discomfort.</p> <p>Use of a knee brace might be recommended initially. This can provide support and stability as you walk.</p> <p>Walking will gradually increase in distance and duration as you progress through physical therapy. It's always best to discuss your recovery timeline, including when you can expect to walk without limitations, with your doctor or physical therapist. They can provide a personalized plan based on your individual circumstances.</p>

## DISCUSSION

The widespread availability of information has significantly impacted the medical field, providing patients with numerous resources to answer their questions. While the internet can serve as a valuable tool for education, it also introduces the risk of disseminating inaccurate or misleading information. AI represents one of the newest developments in information delivery, providing efficient access to knowledge. This study examines the potential role of ChatGPT and Google's Gemini in the medical industry. Existing literature has explored how ChatGPT and Gemini may be utilized to assist in making diagnoses in a multitude of fields, including: ophthalmology, radiology, and emergency medicine (Madadi, Delsoz, Lao, et al. 2023; Delsoz, Madadi, Munir, et al. 2023; Berg, van Bakel, van de Wouw, et al. 2023; Srivastav, Chandrakar, Gupta, et al. 2023). The systems have also demonstrated their capabilities when com-

pleting multiple choice style board examination questions across multiple specialty exams (Madadi, Delsoz, Lao, et al. 2023; Delsoz, Madadi, Munir, et al. 2023; Kuroiwa, Sarcon, Ibara, et al. 2023; Bahir, Zur, Attal, et al. 2024).

The scores varied between surgeons with ChatGPT and Gemini with respect to ACLR related questions. ChatGPT outperformed Gemini on 9 of the 13 questions on average. The scores ranged from 1 (excellent response) to 4 (requiring major clarification). The surgeon evaluators felt ChatGPT responses required less clarification, representing clear and accurate responses. Variability was noted in the responses from surgeons on the same questions, as well as across the entire question set for ChatGPT responses. This variability was only shown to be significant between surgeons across the entire question set when analyzing Gemini responses. ChatGPT's scores varied significantly both between different questions and between surgeons, suggesting its performance was influenced by the specific question content and individual surgeon preferences. These find-

Surgeon	System	Average Score	Number of Questions Scored as Useful
1	ChatGPT	1.85	11
	Gemini	1.62	10
2	ChatGPT	1.77	12
	Gemini	1.77	12
3	ChatGPT	1.15	13
	Gemini	1.31	13
4	ChatGPT	1.69	12
	Gemini	2.00	13
5	ChatGPT	2.23**	12
	Gemini	3.15	9
Average	ChatGPT	1.74	12
	Gemini	1.97	11.4

\*\*Indicates Statistical significance  $p < 0.05$

**Table 3. Average Scores Across Surgeons Between Systems**

ings show that while both systems had variations in how their responses were scored, the patterns of these variations were different.

In addition to quantitative scoring, surgeons evaluated the utility of each response using a binary “yes” or “no.” ChatGPT’s responses were deemed useful an average of 12 times per surgeon, compared to Gemini, which was rated useful an average of 11.4 times. While ChatGPT may have provided technically superior answers to specific questions, Gemini still formulated responses that were perceived as helpful to patients.

The findings demonstrate AI’s ability to provide answers to questions that may be proposed by patients. As such, these new information resources may become useful tools for patients to learn about their orthopaedic procedure and the recovery associated with it. Alternatively, chat-based systems may provide false information. A proposed reason for providing false information is that some large language models (LLM), including ChatGPT have knowledge cut-off dates (OpenAI 2023a). Specifically ChatGPT’s cut-off date is October 2023, indicating that much of its knowledge base stems from prior to that date (OpenAI 2023a). While ChatGPT indicates to its user that it is able to search the web despite its knowledge cut-off, it is still possible the system may lack up to date resources. In contrast, Google’s Gemini does not possess a quoted cut-off date, representing a potential advantage over ChatGPT, however this is unclear.

While this current study sought to evaluate the ability of ChatGPT and Gemini to answer questions related to a procedure or diagnosis, literature exists that aimed to evaluate whether patients can reliably use AI systems to self-diagnose common orthopaedic conditions (Kuroiwa, Sarcon, Ibara, et al. 2023). Kuroiwa et al. found ChatGPT and Gemini could provide answers to the medical questions posed,

however, their accuracy and reliability were inconsistent (Kuroiwa, Sarcon, Ibara, et al. 2023). In this study, ChatGPT outperformed Gemini on 9 of 13 questions and received higher usefulness ratings from surgeons, averaging 12 “yes” responses compared to Gemini’s 11.4. The variability in surgeon evaluations highlights how individual interpretation impacts the assessment of AI responses. By including disclaimers urging users to consult healthcare professionals, both systems acknowledge these limitations and emphasize the need for expert oversight when using AI-generated medical information.

Readability analysis further distinguished the systems. Gemini obtained a higher FK Reading Ease score and a lower FK Grade Level score (11.06) compared to ChatGPT (12.00), indicating that Gemini’s responses were easier to read and required about one less year of education to understand. However, both systems presented information at a level far above the average American reading level of 7th/8th grade (Marchand 2017). The readability scores reinforce the notion that information provided by these LLMs, while useful to the appropriate audience, may be too complex for the average American to understand. This heightens the concern that the information patients access may be misunderstood or misinterpreted if it cannot fully be understood.

Further concern exists with respect to the use of LLMs and personal health information. ChatGPT states it can delete data however, it is not able to delete the question prompts (OpenAI 2023b). Open AI explicitly warns not to include any personal information in prompts (OpenAI 2023b). Additionally, Gemini retains your personal information even if you delete your account for “as long as reasonably necessary to provide services to you.” (Gemini 2024) Although medical history is not explicitly stated as

the type of personal information they save, caution should be exercised. Patients may decide to include personal health details in their queries with ChatGPT or Gemini as a means to receive the most accurate response. This practice carries the risk of exposing sensitive health information due to the uncertain privacy safeguards of AI systems. It also raises ethical concerns regarding the use of user inputs to enhance the functionality of these systems.

Our investigation highlights that ChatGPT and Gemini can provide answers with relative accuracy to patients, however, this does not take into account a major factor: surgeon to surgeon differences. For instance it was found that operative time may vary based on facility and based on each specific surgeon (Glance et al. 2018; Strum et al. 2000). When looking at ACLR specifically, there are a multitude of different rehabilitation plans, each with their own specific phases (Nelson et al. 2021). Other factors can complicate rehabilitation as well such as whether the procedure is a primary repair or revision procedure. Revision procedures present an area for variation in protocol such as longer time in a brace and longer time to return to sport (Rugg et al. 2020).

Such variation reinforces the notion that specifics should still come from the surgeon or healthcare provider. This was reflected in the answers provided by ChatGPT and Gemini all containing some level of disclaimer. A disclaimer is reassuring, however, whether a patient will decide to reach out to their healthcare provider, or utilize the information provided by AI technology remains unknown. This then has the potential to introduce discordance into the relationship between the orthopaedic surgeon and patient if expectations are not met. This relationship is crucial to uphold due to the strong implications that expectations play on patient success and satisfaction. Conner-Spady et al. found that in total hip replacement (THR) and total knee replacement (TKR) patients, they were significantly more satisfied when expectations were met (Conner-Spady et al. 2020). This highlights how crucial it is for patients to address expectations with their surgeon prior to their procedure, not just AI sources.

Further studies should be performed to learn how patients use ChatGPT and Gemini to answer medical questions and how they then use this information. In addition, how physicians view AI resources and how they feel it af-

fects the patient-doctor relationship needs to be investigated.

#### LIMITATIONS

This study was not devoid of limitations. The first major limitation was the fact that questions were formulated using research regarding patient expectations as opposed to information directly from patients. While surveying patients directly would provide more accurate questions that could be posed to ChatGPT, this would be unlikely to substantially change the response from the AI system. The other limitation is the fact that only one orthopaedic surgeon was involved in the process of validating patient questions and validity of the data collected from both ChatGPT and the literature search. In the future, a larger set of questions and more experts in the field would help to explore this question further.

#### CONCLUSION

ChatGPT and Gemini, among other similar artificial intelligence systems, have both garnered interest with incredible speed among medical professionals. These results clearly elucidate how ChatGPT and Gemini are viable resources for patients to obtain general information, however, this information lacks crucially specific details. While these findings are promising, they also underscore the importance of continued evaluation and refinement of AI models for medical use, particularly to ensure that they meet the diverse needs of healthcare professionals and patients. No orthopaedic procedure is exactly the same and no patient is exactly the same, therefore, patients should continue to consult their surgeon for specific details about their injury, surgery, and rehabilitation. While patients may continue to turn to online resources, it is important for the orthopaedic surgeon to understand patient expectations, whether they come from AI resources or not, as to not place strain on the surgeon-patient relationship. As the medical field adapts to these new technological advances, more work is needed to determine what role artificial intelligence may play and what impact it can have on patients and surgeons alike.

Submitted: July 19, 2025 EDT. Accepted: October 08, 2025 EDT. Published: March 01, 2026 EDT.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-ND-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc-nd/4.0> and legal code at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> for more information.

## REFERENCES

- Ayers, J. W., A. Poliak, M. Dredze, et al. 2023. "Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum." *JAMA Intern Med* 183 (6): 589–96. <https://doi.org/10.1001/jamainternmed.2023.1838>.
- Bahir, D., O. Zur, L. Attal, et al. 2024. "Gemini AI vs. ChatGPT: A Comprehensive Examination alongside Ophthalmology Residents in Medical Knowledge." *Graefes Arch Clin Exp Ophthalmol Albrecht Von Graefes Arch Klin Exp Ophthalmol*, September. <https://doi.org/10.1007/s00417-024-06625-4>.
- Berg, H. T., B. van Bakel, L. van de Wouw, et al. 2023. "ChatGPT and Generating a Differential Diagnosis Early in an Emergency Department Presentation." *Ann Emerg Med*, September. <https://doi.org/10.1016/j.annemergmed.2023.08.003>.
- Cascella, M., J. Montomoli, V. Bellini, and E. Bignami. 2023. "Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios." *J Med Syst* 47 (1): 33. <https://doi.org/10.1007/s10916-023-01925-4>.
- Conner-Spady, B. L., E. Bohm, L. Loucks, M. J. Dunbar, D. A. Marshall, and T. W. Noseworthy. 2020. "Patient Expectations and Satisfaction 6 and 12 Months Following Total Hip and Knee Replacement." *Qual Life Res* 29 (3): 705–19. <https://doi.org/10.1007/s11136-019-02359-7>.
- Delsoz, M., Y. Madadi, W. M. Munir, et al. 2023. "Performance of ChatGPT in Diagnosis of Corneal Eye Diseases." *MedRxiv Prepr Serv Health Sci*, August. <https://doi.org/10.1101/2023.08.25.23294635>.
- Gemini. 2024. "Privacy Policy." Gemini. February 26, 2024. <https://www.gemini.com/legal/privacy-policy#section-6-childrens-personal-information>.
- Gilson, A., C. W. Safranek, T. Huang, et al. 2023. "How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment." *JMIR Med Educ* 9:e45312. <https://doi.org/10.2196/45312>.
- Glance, L. G., R. P. Dutton, C. Feng, Y. Li, S. J. Lustik, and A. W. Dick. 2018. "Variability in Case Durations for Common Surgical Procedures." *Anesth Analg* 126 (6): 2017–24. <https://doi.org/10.1213/ANE.0000000000002882>.
- Hasley, H. L., L. G. Bukowiec, J. M. Zaifman, et al. 2023. "Cross-Platform Social Media Analysis Regarding ACL Injury and Surgery." *Curr Orthop Pract* 34 (6): 296–301. <https://doi.org/10.1097/BCO.0000000000001229>.
- Jia, X., Y. Pang, and L. S. Liu. 2021. "Online Health Information Seeking Behavior: A Systematic Review." *Healthc Basel Switz* 9 (12). <https://doi.org/10.3390/healthcare9121740>.
- Johns, W. L., B. J. Martinazzi, B. Miltenberg, H. H. Nam, and S. Hammoud. 2024. "ChatGPT Provides Unsatisfactory Responses to Frequently Asked Questions Regarding Anterior Cruciate Ligament Reconstruction." *Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc* 40 (7): 2067–2079.e1. <https://doi.org/10.1016/j.arthro.2024.01.017>.
- Johnson, D., R. Goodman, J. Patrinely, et al. 2023. "Assessing the Accuracy and Reliability of AI-Generated Medical Responses: An Evaluation of the Chat-GPT Model." <https://doi.org/10.21203/rs.3.rs-2566942/v1>.
- Kung, T. H., M. Cheatham, A. Medenilla, et al. 2023. "Performance of ChatGPT on USMLE: Potential for AI-Assisted Medical Education Using Large Language Models." Edited by A. Dagan. *PLOS Digit Health* 2 (2): e0000198. <https://doi.org/10.1371/journal.pdig.0000198>.
- Kuroiwa, T., A. Sarcon, T. Ibara, et al. 2023. "The Potential of ChatGPT as a Self-Diagnostic Tool in Common Orthopedic Diseases: Exploratory Study." *J Med Internet Res* 25:e47621. <https://doi.org/10.2196/47621>.
- LaValley, S. A., M. T. Kiviniemi, and E. A. Gage-Bouchard. 2017. "Where People Look for Online Health Information." *Health Inf Libr J* 34 (2): 146–55. <https://doi.org/10.1111/hir.12143>.
- Leathers, M. P., A. Merz, J. Wong, T. Scott, J. C. Wang, and S. L. Hame. 2015. "Trends and Demographics in Anterior Cruciate Ligament Reconstruction in the United States." *J Knee Surg* 28 (5): 390–94. <https://doi.org/10.1055/s-0035-1544193>.
- Li, L. T., M. A. Sinkler, J. M. Adelstein, J. E. Voos, and J. G. Calcei. 2024. "ChatGPT Responses to Common Questions About Anterior Cruciate Ligament Reconstruction Are Frequently Satisfactory." *Arthrosc J Arthrosc Relat Surg Off Publ Arthrosc Assoc N Am Int Arthrosc Assoc* 40 (7): 2058–66. <https://doi.org/10.1016/j.arthro.2023.12.009>.

Madadi, Y., M. Delsoz, P. A. Lao, et al. 2023. "ChatGPT Assisting Diagnosis of Neuro-Ophthalmology Diseases Based on Case Reports." *MedRxiv Prepr Serv Health Sci*, September. <https://doi.org/10.1101/2023.09.13.23295508>.

Marchand, L. 2017. "What Is Readability and Why Should Content Editors Care about It?" Center for Plain Language. March 22, 2017. <https://centerforplainlanguage.org/what-is-readability/>.

Nelson, C., L. Rajan, J. Day, R. Hinton, and B. M. Bodendorfer. 2021. "Postoperative Rehabilitation of Anterior Cruciate Ligament Reconstruction: A Systematic Review." *Sports Med Arthrosc Rev* 29 (2): 63–80. <https://doi.org/10.1097/JSA.0000000000000314>.

OpenAI. 2023a. "About OpenAI." OpenAI. 2023. <https://openai.com>.

———. 2023b. "What Is ChatGPT?" OpenAI Help Center. 2023. <https://help.openai.com/en/articles/6783457-what-is-chatgpt>.

Pirkle, S., J. Yang, and T. J. Blumberg. 2024. "Do ChatGPT and Gemini Provide Appropriate Recommendations for Pediatric Orthopaedic Conditions?" *J Pediatr Orthop*, August. <https://doi.org/10.1097/BPO.0000000000002797>.

Rugg, C. M., A. A. Pitcher, C. Allen, and N. K. Pandya. 2020. "Revision ACL Reconstruction in Adolescent Patients." *Orthop J Sports Med* 8 (9): 232596712095333. <https://doi.org/10.1177/2325967120953337>.

Srivastav, S., R. Chandrakar, S. Gupta, et al. 2023. "ChatGPT in Radiology: The Advantages and Limitations of Artificial Intelligence for Medical Imaging Diagnosis." *Cureus* 15 (7): e41435. <https://doi.org/10.7759/cureus.41435>.

Strum, D. P., A. R. Sampson, J. H. May, and L. G. Vargas. 2000. "Surgeon and Type of Anesthesia Predict Variability in Surgical Procedure Times." *Anesthesiology* 92 (5): 1454–66. <https://doi.org/10.1097/0000542-200005000-00036>.