Available online at www.medicinescience.org

ORIGINAL ARTICLE

Medicine Science International Medical Journal

Medicine Science 2026;15(4):1570-8

Comparative evaluation of Google, ChatGPT, and Gemini for answering patient questions about robot-assisted total knee arthroplasty: A cross-sectional study

Cihangir Turemis[1], Ramadan Ozmanevra[2], Firat Erpala[1]

[1]Çeşme Alper Çizgenakat State Hospital, Department of Orthopedics and Traumatology, İzmir, Türkiye
[2]Cyprus International, University Department of Orthopedics and Traumatology, Nicosia, Northern Cyprus

Received 19 August 2026; Accepted 23 October 2026
Available online 27 November 2026 with doi: 10.5455/medscience.2026.08.192

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial-NonDerivatives 4.0 International License.

Abstract

Patients facing total knee arthroplasty (TKA) frequently turn to the internet to supplement their preoperative counseling. However, the quality of information varies considerably between traditional search algorithms and generative artificial intelligence. This cross-sectional study evaluated the accuracy, reliability, and source distribution of answers provided by Google Search, ChatGPT-4, and Google Gemini to patient questions about robot-assisted TKA. We collected the 20 most frequent patient questions about robot-assisted TKA from Google Search, ChatGPT-4, and Gemini (n = 60 total queries). Questions were classified using Rothwell's system, and their baseline web sources were categorized as academic, private practice, surgeon blog, government, or commercial sources. Ten board-certified, fellowship-trained joint arthroplasty surgeons graded a total of 600 evaluation entries in a blinded manner on a 1–5 Likert scale across six domains: relevance, accuracy, clarity, completeness, evidence, and consistency, using the 2022 American Academy of Orthopaedic Surgeons (AAOS) Clinical Practice Guidelines as the gold standard. Overall and post hoc differences were analyzed using Kruskal-Wallis and Bonferroni-adjusted Mann-Whitney U tests. Technical details of surgery represented the primary area of patient concern (28.3%). While Google Search directed users predominantly to private medical practices (40%), ChatGPT-4 (50%) and Gemini (60%) relied heavily on peer-reviewed academic institutions. Blinded specialist scoring showed that both large language models (LLMs) significantly outperformed Google Search in relevance, accuracy, clarity, evidence, and consistency ($p < 0.001$). Gemini achieved the highest scores for text clarity ($p < 0.001$). No significant differences were found in completeness ($p = 0.342$) or between the two LLMs' cumulative quality scores (27.00 ± 1.30 vs. 27.01 ± 1.25, $p > 0.05$). Generative AI platforms provide more objective, uniform, and evidence-based information on robot-assisted TKA than traditional search engines, which are heavily biased by commercial optimization. While they cannot replace direct surgical consultation, LLMs serve as excellent educational adjuncts when used under clinician oversight.

Keywords: Large language models; infodemiology; robotic surgical procedures; total knee arthroplasty; artificial intelligence; ChatGPT

Introduction

Total knee arthroplasty (TKA) remains one of the most successful interventions in modern orthopedics for resolving severe knee osteoarthritis [1]. Even with refined protocols, 15% to 20% of patients experience chronic dissatisfaction after routine manual knee replacements [2,3]. These poor outcomes often result from minor component malalignment, subtle ligamentous imbalances, or rotational outliers. To bypass these technical limits, robot-assisted TKA configurations (such as MAKO, ROSA, and NAVIO) have been widely integrated into modern orthopedic departments [4-6]. These computer-guided modalities utilize precise three dimension (3D) bone mapping, live soft-tissue kinetic balance data, and physical haptic barriers to optimize component positioning [7]. Comparative trials confirm that robotic systems cut down on frontal or sagittal alignment outliers while protecting the surrounding periarticular soft-tissue sleeve.

Due to the technical complexity and novelty of these robotic systems, patients are actively driven to seek supplemental education online; consequently, the spread of digital medical networks has fundamentally altered how patients evaluate their surgical options [8]. Surveys indicate that over 70% of individuals browse the web to learn about operative indications, potential

CITATION
Turemis C, Ozmanevra R, Erpala F. Comparative evaluation of Google, ChatGPT, and Gemini for answering patient questions about robot-assisted total knee arthroplasty: A cross-sectional study. Med Science. 2026;15(4):1570-8.

Corresponding Author: Cihangir Turemis, Çeşme Alper Çizgenakat State Hospital, Department of Orthopedics and Traumatology, İzmir, Türkiye
Email: cihangirturemis @gmail.com

DOI: 10.5455/medscience.2026.08.192

complications, and physical rehabilitation intervals long before scheduling an operation [9]. Though open access to medical facts encourages patient autonomy, routine web searches frequently yield unverified marketing material, low-yield clinical blogs, or biased patient forums. This unchecked mix of commercial and clinical text introduces unnecessary patient anxiety, fosters distorted expectations, and strains the direct physician-patient relationship.

Recently, natural language processing tools—specifically OpenAI's ChatGPT-4 and Google AI's Gemini—have shifted the landscape of online data retrieval [10]. Traditional search engines rank indexed hyperlinks according to optimization tactics (SEO) and commercial bidding. In contrast, large language models (LLMs) synthetically build unified, conversational paragraphs to address specific prompts [11]. Standard performance testing shows that these conversational tools answer general clinical inquiries with a high degree of fidelity across various medical specialties.

However, incorporating automated language models into high-stakes surgical subspecialties like adult joint reconstruction requires close inspection. Because these models predict subsequent words based on mathematical probabilities, they can present fabricated facts or clinical errors with high structural confidence—a vulnerability termed medical hallucination. For example, an AI model might confidently, yet incorrectly, invent a non-existent surgical complication or falsely claim that a robotic procedure requires no postoperative physical therapy. Given the novelty, the required learning curve, and the significant expenses associated with robotic knee surgeries, it is imperative to assess the accuracy of the medical information these platforms deliver to patients.

This cross-sectional study systematically compared the quality, objective clinical accuracy, and source distribution of patient-directed explanations regarding robotic TKA across Google Search, ChatGPT-4, and Google Gemini. Additionally, a secondary aim was to explicitly analyze the source distribution to determine where these distinct digital platforms derive their medical data. We hypothesized that the generative platforms would achieve superior objective marks by selecting peer-reviewed academic frameworks over commercially optimized index pages.

Material and Methods

Study Design

This study was designed as a cross-sectional, observational analysis to evaluate the quality and reliability of artificial intelligence platforms compared to search engine algorithms in answering patient inquiries about robot-assisted knee replacement surgery. Because this study was based entirely on publicly available, anonymous, and publicly accessible web content, ethics committee approval and informed consent were not required. This cross-sectional study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines.

To collect clean, non-personalized text uncorrupted by targeted tracking loops, all search actions were conducted by a single researcher on clean web browser configurations utilizing incognito tabs. The researcher logged into newly created, dedicated research accounts with no prior search or chat history. Geolocation data, search history files, tracking cookies, and local browser cache directories were completely purged prior to the initialization of the study. The chosen clinical target strings were defined as "Robot-Assisted Knee Replacement Surgery" and "Robotic Knee Surgery". All search queries and data extraction across the three platforms (Google Search, ChatGPT-4 March 2026 version, and Google Gemini March 2026 version) were executed between March and April 2026.

The 20 most frequent user questions were extracted from Google's automated algorithmic dropdowns ("People Also Ask"). Symmetrically, identical semantic text entries were typed into non-personalized ChatGPT-4 and Google Gemini interfaces to isolate the 20 most prevalent questions consumers ask about robotic joint procedures. Irrelevant entries or non-specific general joint questions (e.g., "What is severe gonarthrosis?") were removed from the final dataset. The 60 isolated inquiries are displayed within the technical data matrices of Table 1, Table 2, and Table 3.

Taxonomic and Sourced Asset Classification

Two joint reconstruction specialists independently sorted all 60 unique questions using the Rothwell health data framework. This taxonomy groups patient inquiries into distinct semantic pockets: Indications/Management, Technical Details, Risks/Complications, Restrictions, Timeline of Recovery, Pain, Advantage, Safety, and Cost. Divergences in classification were settled through open discussion involving a third senior joint reconstructive reader. Concurrently, the digital footprints of the background references were tracked and classified into five verified domain types: Academic/Institutional, Private Medical Practice, Single Surgeon Blog, Government/Public Health, and Commercial Organizations.

Blinded Specialist Evaluation and Scoring

The grading board consisted of ten (n = 10) independent, board-certified orthopedic surgeons with at least three years of post-fellowship experience in primary and revision joint arthroplasty. To prevent brand alignment or confirmation bias, the raw text files of all 60 answers were stripped of web formatting, layout configurations, original font settings, hyperlinks, and explicit platform names. The normalized text blocks were shuffled and placed onto a randomized digital evaluation panel. All surgeons remained entirely blinded to the source platform of any given response.

To align scoring criteria, all specialists reviewed the 2022 American Academy of Orthopaedic Surgeons (AAOS) [12] Clinical Practice Guidelines for the Surgical Treatment of Knee Osteoarthritis before beginning evaluation. Each answer text was graded across six distinct clinical parameters using a 5-point Likert scale.

DOI: 10.5455/medscience.2026.08.192

Table 1. Detailed patient-led questions and generated contextual outputs from the ChatGPT 4 platform

Subcategory matrix	Sample question documented	Standardized core synthesized output context
Indication/management	What is robotic knee arthroplasty, and how does it work?	Robotic knee arthroplasty uses robotic systems to assist surgeons in performing knee replacement surgeries with enhanced precision and accuracy.
Indication/management	Who is a good candidate for robotic-assisted knee replacement?	Patients with severe arthritis or joint damage who are suitable for knee replacement surgery may benefit, depending on their health and anatomy.
Indication/management	What are the benefits of robotic knee surgery?	Benefits include improved accuracy, reduced post-operative pain, faster recovery, and possibly longer-lasting implants.
Technical details	What types of knee replacements can be done robotically?	Both partial and total knee replacements can be performed using robotic systems.
Technical details	Does the robot perform the surgery?	No, the surgeon controls the procedure while the robot provides guidance and precision.
Technical details	What robotic systems are commonly used for knee replacement?	Popular systems include MAKO, ROSA, Navio, and VELYS.
Technical details	What kind of imaging is required for robotic knee surgery?	Preoperative imaging like CT scans or MRIs is often needed to create a detailed 3D model of the knee.
Technical details	How long does robotic knee replacement surgery take?	It may take slightly longer than traditional surgery due to additional preoperative planning and robotic system setup.
Risks/complications	What happens if the robotic system fails during surgery?	Surgeons are fully trained to continue the procedure using traditional techniques if needed.
Risks/complications	Are there additional risks with robotic-assisted surgery?	While the surgery is generally safe, risks include potential technical issues and standard surgical risks like infection or blood clots.
Risks/complications	Does robotic surgery reduce complications?	Precision can decrease complications like implant misalignment, which may lead to better outcomes.
Risks/complications	Is robotic surgery safe for older adults?	Yes, but a patient's overall health and surgical risk factors need to be assessed.
Timeline of recovery	What is the recovery time after robotic knee replacement?	Recovery varies but often ranges from 4–6 weeks for basic mobility and 3–6 months for full recovery.
Timeline of recovery	How long will my knee replacement last?	Robotic precision may lead to longer-lasting implants, potentially 15–20 years or more.
Pain	Does robotic surgery result in less postoperative pain?	Many patients report less pain and quicker recovery due to improved surgical accuracy and minimal soft tissue damage.
Advantage	How does robotic knee surgery differ from traditional knee replacement?	It offers improved precision in implant placement, better alignment, and potentially shorter recovery times.
Safety	Is robotic knee surgery FDA approved?	Yes, robotic systems used in knee replacement are FDA-approved and have undergone rigorous testing.
Safety	How precise is the robotic system?	The system allows for millimeter-level precision in bone cutting and implant placement.
Cost	Is robotic knee replacement more expensive?	It can be, due to the use of advanced technology, but insurance usually covers medically necessary procedures.
Cost	Does insurance cover robotic assisted knee replacement?	Most insurance plans, including Medicare, cover robotic surgery if deemed medically necessary.

Abbreviations: 3D, three-dimensional; CT, computed tomography; FDA, Food and Drug Administration; MRI, magnetic resonance imaging.

DOI: 10.5455/medscience.2026.08.192

Table 2. Detailed patient-led questions and extracted core content from the Google search engine

Subcategory matrix	Sample question documented	Standardized core synthesized output context
Indication/management	Who is not a candidate for robotic knee replacement?	Certain joint and bone deformities can prevent an individual from having robotic knee replacement. This can include health conditions like advanced osteoporosis, rheumatoid arthritis and bow legs or knock knees.
Indication/management	How many stitches are in a robotic knee replacement?	The incision is then closed with stitches that are all buried underneath the skin and dissolve on their own over time. There are no stitches or staples that need to be removed.
Indication/management	What are the disadvantages of robotic TKA?	Its high costs, reliance on technology, and lack of long-term evidence are some limitations.
Technical details	What is the downside to robotic knee replacement?	However, the main disadvantages of robot assisted knee surgery are a slightly longer surgery time and a possibility for an additional, very small incision for pin placement.
Technical details	What is the difference between a knee replacement and a robotic knee replacement?	A robotic knee replacement is similar to a traditional knee replacement. Your surgeon removes damaged tissue in your knee and replaces it with an artificial joint.
Technical details	Are you awake during robotic knee replacement?	You will not be awake during surgery. Spinal anesthesia provides surgical anesthesia and you will be given sedatives to help you relax and put you in a light sleep.
Technical details	What is the newest knee replacement option?	In minimally invasive total knee replacement: The surgical procedure is similar to a traditional total knee replacement, but there is less cutting of the tissue surrounding the knee.
Technical details	How big is the incision for robotic knee replacement?	Minimally invasive surgery for knee replacement involves the use of smaller incisions which are only 4 to 6 inches in length as compared to the 10-12 inch long incision used in the traditional procedure.
Technical details	What is the newest procedure for knee replacement?	Minimally-invasive quadriceps-sparing total knee replacement is a new surgical technique that allows surgeons to insert the same time-tested components.
Technical details	How long do you stay in the hospital after robotic knee replacement surgery?	Hospital stays can average anywhere from one to three days.
Risks/complications	What is the success rate of robotic knee replacement?	On average, the success rate of a knee replacement surgery is about 95%; however, with a robotic knee surgery, there's a significantly reduced complication rate.
Restrictions	How soon can you drive after robotic knee surgery?	If surgery was on your RIGHT knee, you should not drive for 1 month (4 weeks) after surgery.
Restrictions	What can you never do after knee replacement?	You shouldn't downhill ski or play contact sports such as football and soccer. In general, avoid sports that require jerking, twisting, pulling, or running.
Restrictions	What are the mistakes people make after knee surgery?	Top five mistakes after a knee replacement: doing too much too soon, skipping your physio routine, ignoring severe pain or complications, forgetting about wound care and medication.
Timeline of recovery	What is the recovery time for robotic assisted knee replacement?	Most people are walking unassisted by the second or third week after surgery. You'll go through at least six weeks of physical therapy.
Timeline of recovery	Do you need physical therapy after robotic knee replacement?	Physical therapy is critical to regain strength and motion. Skipping this introduces a high risk of permanent joint stiffness or construct harm.
Advantage	How painful is robotic knee surgery?	Since the robotic arm operates with pinpoint accuracy, the incision is much smaller and more cosmetic. Thus, it allows postoperative maintenance of muscle function.
Advantage	How much bed rest is required after knee surgery?	The duration of bed rest after knee replacement surgery is relatively short. Most patients are encouraged to get up and start moving as soon as possible after the procedure.
Cost	What is the average cost of robotic knee replacement?	Total knee replacement: $32,500; Partial knee replacement: $11,350 to $39,213; Robotic-assisted knee replacement: $15,607 to $53,918.
Cost	Does insurance cover robotic knee surgery?	Any insurance that covers minimally invasive surgery generally covers Robotic Surgery if deemed medically necessary.

Abbreviation: TKA, total knee arthroplasty.

DOI: 10.5455/medscience.2026.08.192

Table 3. Detailed patient-led questions and generated contextual outputs from the Google Gemini AI platform

Subcategory matrix	Sample question documented	Standardized core synthesized output context
Indication/management	Is robotic knee replacement suitable for severe bone-on-bone osteoarthritis?	Yes, robotic systems are highly effective for end-stage (bone-on-bone) osteoarthritis, as they allow customized cutting blocks to correct severe deformities.
Indication/management	Can a robotic knee replacement be done as an outpatient procedure?	Depending on patient comorbidities and institutional criteria, robotic TKA can facilitate same-day discharge due to reduced soft-tissue trauma.
Indication/management	What are the long-term clinical trial outcomes for robotic knee replacements?	Mid-term registries show excellent survival rates and fewer early revisions; however, 20-year survival data are still being compiled compared to manual TKA.
Technical details	How does the surgeon map the knee anatomy during a robotic procedure?	The system utilizes optical trackers, intraoperative registration pins, or preoperative 3D-CT scans to create a real-time kinematic map of the joint.
Technical details	What is the role of artificial intelligence in robotic TKA?	AI algorithms process intraoperative ligament balance data and calculate the optimal implant size, rotation, and translation before any bone cuts are made.
Technical details	Does the robotic arm move independently during the surgery?	No, the robotic arm operates in a semi-active mode; it remains completely under the physical control of the surgeon and freezes automatically if boundaries are breached.
Technical details	Are custom-made 3D implants used in robotic knee surgery?	No, standard time-tested implants are used, but the robot customizes their exact spatial position, alignment, and tilt to match the patient's unique anatomy.
Technical details	How deep do the registration pins go into the bone?	The optical tracking arrays require small pins drilled into the femur and tibia, which heal completely within a few weeks postoperatively.
Risks/complications	Can the tracking pins cause bone fractures or infections?	Pin-site complications, including local infection or stress fractures, are rare ($ < 1\%$) but represent a unique risk specific to robotic systems.
Risks/complications	Is there a risk of the computer software freezing or malfunctioning mid-surgery?	Software outliers are rare; if a technical failure occurs, the system undergoes an automated backup restart, or the surgeon safely transitions to manual instruments.
Risks/complications	Does robotic surgery increase the risk of postoperative blood clots?	No, the risk of deep vein thrombosis (DVT) is comparable to traditional surgery, and standard chemical/mechanical prophylaxis is utilized postoperatively.
Timeline of recovery	When can I bend my knee completely after robotic knee replacement?	Most patients achieve 90° of flexion within the first week, and progress to functional flexion ($110^\circ\text{--}120^\circ$) by week 4 to 6 through physical therapy.
Timeline of recovery	How soon can I return to low-impact sports after robotic TKA?	Patients can generally resume low-impact activities like cycling, swimming, or golfing around 6 to 12 weeks postoperatively, following clearance.
Pain	Why do some patients report less pain with robotic knee surgery?	The robotic boundary control prevents accidental saw blade deviation, significantly reducing thermal and mechanical injury to the collateral ligaments and capsule.
Pain	How is severe postoperative pain managed after robotic knee arthroplasty?	A multimodal analgesia protocol is utilized, combining adductor canal blocks, periarticular local infiltrations, scheduled non-opioids, and targeted opioids.
Advantage	Does robotic surgery improve the rotational alignment of the knee components?	Yes, real-time tracking allows precise execution of the targeted femoral component rotation, decreasing the incidence of patellofemoral maltracking.
Safety	Are there any age limits or contraindications for robotic knee surgery?	There is no absolute age limit; contraindications are identical to manual TKA, such as active joint infection, severe vascular disease, or incompetent extensor mechanisms.
Safety	How does the robot protect the popliteal artery and sciatic nerve?	The surgeon establishes intraoperative haptic boundaries; the saw automatically shuts off if it approaches the posterior soft-tissue envelope.
Cost	Is there an extra out-of-pocket fee for robotic knee surgery?	Some private centers may charge an additional technology fee, but if the procedure is deemed medically necessary, standard hospital networks cover the cost.
Cost	Does public healthcare or Medicare cover the robotic component of the surgery?	Yes, major public healthcare insurance programs and Medicare cover robotic-assisted knee arthroplasty without sub-charging the robotic modality separately.

Abbreviations: 3D, three-dimensional; AI, artificial intelligence; CT, computed tomography; DVT, deep vein thrombosis; TKA, total knee arthroplasty.

DOI: 10.5455/medscience.2026.08.192

Statistical Analysis

Data computations were executed via SPSS version 23.0 (IBM Corp., Armonk, NY, USA). Continuous measurements are presented as means $\pm$ standard deviations (SD). Data normality was checked using the Kolmogorov-Smirnov test. Because the Likert scale grades exhibited an ordinal, non-parametric distribution, overall inter-group variances across the three test platforms were analyzed with the Kruskal-Wallis H-test. For significant parameters, individual pairwise post-hoc tests were completed using the Mann-Whitney U-test with a strict Bonferroni correction applied to prevent Type I error inflation [13]. Categorical data variables (including Rothwell fields and domain sources) were evaluated using the Pearson chi-square (χ^2) test [14]. Statistical significance was established at $p < 0.05$.

Results

The evaluation phase was completed by a panel of ten independent, board-certified orthopedic surgeons, all possessing at least three years of post-fellowship experience in primary and revision joint arthroplasty. The structural mapping of the 60 patient questions revealed no significant differences in thematic distribution among the three search channels ($p = 0.412$, Table 4). Questions centered on Technical Details were the most common across all engines (n = 17, 28.3%). Interestingly, lifestyle restrictions were entirely absent from the AI question sets (0% vs. 15% in Google Search), while queries tracking perioperative pain and safety protocols were more frequently captured by the two language models (Table 4). Tracking the origin of the information revealed that ChatGPT-4 (50%) and Google Gemini (60%) compiled the majority of their text blocks from Academic/Institutional sites (such as the Mayo Clinic, Cleveland Clinic, and HSS). In contrast, standard Google search requests heavily indexed Private Medical Practices (40% vs. 10% for ChatGPT-4 and 5% for Gemini). While the cross-sectional distribution skewed heavily toward institutional pages for the AI tools, the descriptive difference fell short of statistical significance ($p = 0.115$, Table 4).

The surgical board produced 600 unique evaluation entries. The two LLMs (ChatGPT-4 and Google Gemini) earned significantly higher scores than standard Google Search queries across five of the six evaluated parameters (Table 5).Post-hoc pairwise tests confirmed that both ChatGPT-4 and Gemini surpassed Google Search in relevance, clinical accuracy, evidence base, and text consistency ($p < 0.001$). Google Gemini secured the highest scores in structural clarity, outperforming Google Search ($p < 0.001$). No statistically significant difference was detected among the three platforms regarding text completeness ($p = 0.342$, Table 5). Furthermore, ChatGPT-4 and Google Gemini performed symmetrically, with no significant variation in their cumulative total quality scores (27.00 ± 1.30 vs. 27.01 ± 1.25\$, $p > 0.05$).

Table 4. Cross-sectional distribution matrix of primary question subcategories and sourced website categories across three platforms

Baseline grouping domain	Google Search (n = 20)	ChatGPT-4 (n = 20)	Google Gemini (n = 20)	Combined total (n = 60)	p value (χ2)
Rothwell subcategory class					**0.412**
Indications/management	3 (15%)	3 (15%)	3 (15%)	9 (15.0%)	
Technical details	7 (35%)	5 (25%)	5 (25%)	17 (28.3%)	
Risks/complications	1 (5%)	4 (20%)	3 (15%)	8 (13.3%)	
Restrictions	3 (15%)	0 (0%)	0 (0%)	3 (5.0%)	
Timeline of recovery	2 (10%)	2 (10%)	2 (10%)	6 (10.0%)	
Pain	0 (0%)	1 (5%)	2 (10%)	3 (5.0%)	
Advantage	2 (10%)	1 (5%)	1 (5%)	4 (6.7%)	
Safety	0 (0%)	2 (10%)	2 (10%)	4 (6.7%)	
Cost	2 (10%)	2 (10%)	2 (10%)	6 (10.0%)	
Website resource category					**0.115**
Academic / institutional	5 (25%)	10 (50%)	12 (60%)	27 (45.0%)	
Private medical practice	8 (40%)	2 (10%)	1 (5%)	11 (18.3%)	
Single surgeon blog	2 (10%)	2 (10%)	2 (10%)	6 (10.0%)	
Government / public health	0 (0%)	1 (5%)	0 (0%)	1 (1.7%)	
Commercial entities	5 (25%)	5 (25%)	5 (25%)	15 (25.0%)	

DOI: 10.5455/medscience.2026.08.192

Table 5. Blinded orthopedic surgeon evaluation ratings of generated answers across three digital platforms

Evaluated semantic parameter	Google outcomes (n = 200)	ChatGPT-4 outcomes (n = 200)	Google Gemini outcomes (n = 200)	p value (Kruskal-Wallis)	Post-hoc significance
Relevance	3.90 ± 0.60	4.40 ± 0.50	4.45 ± 0.48	< 0.001*	LLMs > Google Search (p < 0.001)
Accuracy	3.80 ± 0.49	4.40 ± 0.50	4.42 ± 0.51	< 0.001*	LLMs > Google Search (p < 0.001)
Clarity	3.87 ± 0.49	4.50 ± 0.50	4.58 ± 0.46	< 0.001*	Google Gemini ≥ ChatGPT > Google Search
Completeness	4.48 ± 0.51	4.50 ± 0.50	4.41 ± 0.53	0.342	No Significant Difference (p > 0.05)
Evidence	4.32 ± 0.56	4.55 ± 0.49	4.62 ± 0.44	< 0.001*	LLMs > Google Search (p < 0.001)
Consistency	4.39 ± 0.48	4.51 ± 0.50	4.53 ± 0.49	0.008*	LLMs > Google Search (p < 0.05)
Cumulative total score	24.80 ± 1.30	27.00 ± 1.30	27.01 ± 1.25	< 0.001*	LLMs > Google Search (p < 0.001)

Abbreviation: LLMs, large language models.

Discussion

The most significant finding of this study is that generative AI platforms (ChatGPT-4 and Google Gemini) provided significantly more accurate, relevant, and evidence-based information regarding robotic TKA compared to traditional Google Search. While traditional search algorithms frequently directed patients to commercially optimized private practice portals, the LLMs predominantly synthesized their responses from peer-reviewed academic frameworks. As robot-assisted TKA becomes a standard offering in modern reconstructive centers, providing clear and accurate preoperative education is vital for aligning patient expectations and organizing home rehabilitation [15,16]. This three-group study shows that ChatGPT-4 and Google Gemini provide more accurate, clear, and evidence-grounded data on robotic knee procedures than traditional Google Search queries [17].

The most meaningful finding of this analysis lies in the resource origin distribution across the different systems [18,19]. Traditional search tools use indexing models governed heavily by Search Engine SEO tactics and corporate intent [20]. This means they frequently elevate private practice portals or commercial vendor pages above non-profit institutional materials. In our study, 40% of standard Google links directed users to private clinical blogs [21,22]. While these portals can offer introductory details, they carry an inherent conflict of interest and may overemphasize the benefits of robotic technology to attract self-paying patients. Conversely, ChatGPT-4 built half of its answers from non-profit academic facilities, and Google Gemini raised that metric to 60%. This variance explains why the blinded surgical board gave both language models higher marks for medical accuracy and scientific grounding.

The clinical utility of generative AI platforms is further supported by their superior performance in text clarity and relevance [23,24]. Standard search engines force patients to parse multiple sponsored links, pop-up ads, and shifting document styles [25,26]. This fragmentation leads to cognitive overload and patient confusion. Language models minimize this friction by synthesizing disparate source facts into a single, cohesive narrative [27,28]. The identical cumulative scores achieved by ChatGPT-4 and Google Gemini suggest that the leading conversational AI architectures have reached a comparable level of maturity regarding medical infodemiology [29,30].

However, the lack of statistical variance in text completeness (p = 0.342) highlights a structural limitation of automated language models. While these tools write fluent, highly organized prose, they generate general summaries that overlook the crucial nuances of an individual patient's anatomy, residual bone stock, or medical comorbidities [31,32]. For instance, precise answers regarding when it is safe to drive after surgery or the long-term wear rates of a specific polyethylene liner require a tailored assessment by the attending surgeon [33,34].

While these findings support using conversational AI as an educational supplement, integrating automated models into clinical practice presents distinct challenges [35]. Joint specialists cannot readily audit the specific neural pathways or training dataset weights used to build an individual paragraph. This lack of transparency maintains concerns over medical hallucinations—where an AI model generates clinically false or fabricated statements with extreme structural certainty [36,37]. While the evaluating surgeons did not detect any overt medical hallucinations within the specific 60 AI-generated answers evaluated in this study, the possibility of sporadic misinformation remains an active risk. Furthermore, utilizing these advanced conversational tools assumes a baseline level of digital literacy, which could widen gaps in health equity [38]. Patients who struggle to access or query conversational interfaces will remain dependent on lower-quality, advertisement-driven online portals.

The primary strength and original contribution of this study lie in its clinically focused methodology. While previous literature

DOI: 10.5455/medscience.2026.08.192

has frequently evaluated AI outputs using automated readability scores or single-reviewer assessments, this study employed a fully blinded, 10-surgeon expert panel utilizing established AAOS clinical guidelines. This approach moves beyond a simple comparison of applications; it provides a comprehensive, peer-level assessment of how artificial intelligence synthesizes complex surgical data compared to commercial search engines, thereby contributing valuable clinical insights for digital patient education in robotic arthroplasty.

This study has several limitations. First, web search analytics and language model parameters shift rapidly, making these observations a cross-sectional snapshot of current outputs. Second, although the evaluation board was completely blinded, individual clinical preferences among the reviewing surgeons could influence subjective marks. Third, the analysis evaluated English-language entries exclusively; hence, the data patterns may not match online outputs in other languages. Additionally, this study exclusively evaluated text-based search outputs, whereas many patients also rely heavily on audio-visual platforms for their preoperative education. Finally, generative AI outputs depend heavily on the exact structure and phrasing of the user's prompt, meaning different query formulations could yield varying responses. Therefore, future research should prospectively evaluate patient comprehension of these AI outputs in real-world clinical settings and assess the utility of multi-modal AI platforms in surgical education.

Conclusion

Compared to traditional Google Search results, ChatGPT-4 and Google Gemini supply more accurate, readable, and evidence-grounded information on robot-assisted TKA by drawing reliably on academic and institutional networks. However, because AI platforms are vulnerable to hallucinations and lack the capacity for personalized clinical judgment, they cannot replace formal surgical evaluations. When monitored under a physician's oversight, these language tools serve as valuable assets for boosting patient education and supporting shared decision-making, provided that surgeons proactively warn patients about the potential risks of AI hallucinations during their initial clinical consultations.

Ethical Approval

Ethical approval was not required for this study because it exclusively involved the analysis of publicly available web content and did not involve human participants, patient records, or identifiable personal information.

Patient Consent

Informed consent was not required because no human participants were recruited or directly involved in the study, and no personal or identifiable patient data were collected or analyzed.

Conflict of Interest

The authors confirm the absence of any conflicts associated with this work.

Funding

The authors declare that they have received no financial support for the study.

References

1. Gill GS, Joshi AB, Mills DM. Total condylar knee arthroplasty. 16- to 21-year results. Clin Orthop Relat Res. 1999;210-5.
2. Baker PN, van der Meulen JH, Lewsey J, et al. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. J Bone Joint Surg Br. 2007;89:893-900.
3. Parratte S, Pagnano MW, Trousdale RT, Berry DJ. Effect of postoperative mechanical axis alignment on the fifteen-year survival of modern, cemented total knee replacements. J Bone Joint Surg Am. 2010;92:2143-9.
4. Song EK, Seon JK, Yim JH, et al. Robotic-assisted TKA reduces postoperative alignment outliers and improves gap balance compared to conventional TKA. Clin Orthop Relat Res. 2013;471:118-26. Erratum in: Clin Orthop Relat Res. 2012;470:2627.
5. Khlopas A, Chughtai M, Hampp EL, et al. Robotic-arm assisted total knee arthroplasty demonstrated soft tissue protection. Surg Technol Int. 2017;30:441-6.
6. Kayani B, Konan S, Pietrzak JRT, Haddad FS. Iatrogenic bone and soft tissue trauma in robotic-arm assisted total knee arthroplasty compared with conventional jig-based total knee arthroplasty: a prospective cohort study and validation of a new classification system. J Arthroplasty. 2018;33:2496-501.
7. Song EK, Seon JK, Park SJ, et al. Simultaneous bilateral total knee arthroplasty with robotic and conventional techniques: a prospective, randomized study. Knee Surg Sports Traumatol Arthrosc. 2011;19:1069-76.
8. Wen Z, Huang H. The potential for artificial intelligence in healthcare. J Commer Biotechnol. 2022;27:217-24.
9. Rennis L, McNamara G, Seidel E, Shneyderman Y. Google it!: urban community college students' use of the internet to obtain self-care and personal health information. Coll Stud J. 2015;49:414-26.
10. Tharakan S, Klein B, Bartlett L, et al. 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.
11. Rao AJ, Dy CJ, Goldfarb CA, et al. Patient preferences and utilization of online resources for patients treated in hand surgery practices. Hand (N Y). 2019;14:277-83.
12. American Academy of Orthopaedic Surgeons: Surgical Management of Osteoarthritis of the Knee: Evidence-Based Clinical Practice Guideline. https://www.orthoguidelines.org/guidelines access date 18.03.2026
13. Abbas MJ, Khalil LS, Haikal A, et al. Eliciting emotion and action increases social media engagement: an analysis of influential orthopaedic surgeons. Arthrosc Sports Med Rehabil. 2021;3:e1301-8.
14. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns. 2013;92:426-31.
15. Rothwell DJ. SNOMED-based knowledge representation. Methods Inf Med. 1995;34:209-13.
16. Finney Rutten LJ, Blake KD, Greenberg-Worisek AJ, et al. Online health information seeking among US adults: measuring progress toward a Healthy People 2020 objective. Public Health Rep. 2019;134:617-25.
17. Shen SA, Perez-Heydrich CA, Xie DX, Nellis JC. ChatGPT vs. web search for patient questions: what does ChatGPT do better? Eur Arch Otorhinolaryngol. 2024;281:3219-25.

DOI: 10.5455/medscience.2026.08.192

18. O'Mathúna DP. How should clinicians engage with online health information? AMA J Ethics. 2018;20:E1059-66.

19. Gomez-Cabello CA, Borna S, Pressman S, et al. Artificial-intelligence-based clinical decision support systems in primary care: a scoping review of current clinical implementations. Eur J Investig Health Psychol Educ. 2024;14:685-98.

20. Mazurowski MA, Buda M, Saha A, Bashir MR. Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. J Magn Reson Imaging. 2019;49:939-54.

21. Baum ZJ, Yu X, Ayala PY, et al. Artificial intelligence in chemistry: current trends and future directions. J Chem Inf Model. 2021;61:3197-212.

22. Liu H, Ding N, Li X, et al. Artificial intelligence and radiologist burnout. JAMA Netw Open. 2024;7:e2448714.

23. Sendekie AK, Limenh LW, Abate BB, et al. Artificial intelligence in community pharmacy practice: pharmacists' perceptions, willingness to utilize, and barriers to implementation. Explor Res Clin Soc Pharm. 2024;16:100542.

24. The AI revolution is already here: Transforming orthopedics in 2024 and beyond. https://www.sportsmed.org/membership/sports-medicine-update/fall-2024/the-ai-revolution-is-already-here-transforming-orthopedics-in-2024-and-beyond access date 15.03.2026

25. Gala D, Makaryus AN. The utility of language models in cardiology: a narrative review of the benefits and concerns of ChatGPT-4. Int J Environ Res Public Health. 2023;20:6438.

26. Daungsupawong H, Wiwanitkit V. Fertility preservation written by ChatGPT-4: correspondence. J Assist Reprod Genet. 2024;41:1933.

27. Radulesco T, Saibene AM, Michel J, et al. ChatGPT-4 performance in rhinology: a clinical case series. Int Forum Allergy Rhinol. 2024;14:1123-30.

28. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLoS Digit Health. 2023;2:e0000198.

29. Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of value for cardiovascular nurses and allied health professionals. Eur J Cardiovasc Nurs. 2023;22:e55-9.

30. Zhang C, Liu S, Zhou X, et al. Examining the role of large language models in orthopedics: systematic review. J Med Internet Res. 2024;26:e59607.

31. Oeding JF, Lu AZ, Mazzucco M, 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.

32. Kunze KN, Nwachukwu BU, Cote MP, Ramkumar PN. Large language models applied to health care tasks may improve clinical efficiency, value of care rendered, research, and medical education. Arthroscopy. 2025;41:547-56.

33. Sinkler MA, Li LT, Adelstein JM, Strony JT. Editorial commentary: ChatGPT has the potential to be an important patient education tool and may outperform Google. Arthroscopy. 2025;41:598-9.

34. Gao L, Mu H, Lin Y, et al. Review of the current situation of postoperative pain and causes of inadequate pain management in Africa. J Pain Res. 2023;16:1767-78.

35. Nasir M, Ahmed A. Knowledge about postoperative pain and its management in surgical patients. Cureus. 2020;12:e6685.

36. Obana KK, Lind DRG, Mastroianni MA, et al. What are our patients asking Google about acromioclavicular joint injuries?—frequently asked online questions and the quality of online resources. JSES Rev Rep Tech. 2024;4:175-81.

37. Dubin JA, Bains SS, Chen Z, et al. Using a Google web search analysis to assess the utility of ChatGPT in total joint arthroplasty. J Arthroplasty. 2023;38:1195-202.

38. Castle JP, Khalil LS, Tramer JS, et al. Indications for surgery, activities after surgery, and pain are the most commonly asked questions in anterior cruciate ligament injury and reconstruction. Arthrosc Sports Med Rehabil. 2023;5:100805.