

# Performance of ChatGPT-4 for Providing Patients with Medical Information Regarding Total Shoulder Arthroplasty: A Benchmark Analysis against Google Web Search

Michael Robert Mazzucco, Amy Lu<sup>1</sup>, Michael Fu, Ayoosh Pareek, David M Dines<sup>2</sup>, Russell F Warren<sup>3</sup>, Lawrence V Gulotta<sup>4</sup>, Joshua S Dines<sup>2</sup>, Kyle Kunze

<sup>1</sup>Weill Cornell Medicine, <sup>2</sup>Hospital For Special Surgery, <sup>3</sup>Hosp for Special Surgery, <sup>4</sup>Hosp for Special Surg-Cornell

**INTRODUCTION:** The recent introduction and deployment of large language models (LLMs) have demonstrated far-reaching applications capable of expediting infrastructural processes and unlocking novel methods for task completion within health care and many commercial and private sectors. Chat-Generative Pre-trained Transformer (GPT) is an AI chatbot that has garnered a considerable amount of interest for the successful performance of such tasks and dynamic ability to adapt and learn. Given the potential implications that utilizing ChatGPT-4 as a source of patient information for frequently asked questions or medical advice concerning musculoskeletal conditions may confer on reducing provider burden and better informing patients about their current health state, it is imperative to assess the potential utility of ChatGPT-4 such that it can be advocated for or cautioned against for patient use. The purpose of the current study was to investigate the ability of ChatGPT to provide appropriate medical information regarding total shoulder arthroplasty (TSA).

## METHODS:

Two search engines were utilized to perform simulated patient queries: 1) Google Web Search and 2) ChatGPT-4. Google Web Search was performed utilizing a clean-installed Google Chrome Browser on April 25, 2023. A clean-installed browser was utilized in order to minimize contributions of individualized search algorithms and bias potentially present in sponsored advertisements, cookies, cache, and browsing history. Google was chosen as the default search engine for comparison as it is the most widely used search engine worldwide and the only search engine that generates FAQs when prompted with a query, therefore allowing comparisons through a systematic approach. The query “total shoulder replacement” was performed using Google Web Search, and the first 10 frequently asked questions (FAQs) and associated sources were extracted. Utilizing ChatGPT, the following command was then input: “Perform a google search with the search term “total shoulder replacement and record the first 10 FAQs related to the search term.” This search was performed again to identify the first 10 FAQs including numerical responses, after which ChatGPT was prompted with these questions and answers were recorded and compared. The number of academic sources used by Google and ChatGPT-4 were compared using Chi-squared and Fisher’s exact test depending on cell event frequency where appropriate.

## RESULTS:

Concerning FAQs with numeric responses, 8/10 (80%) had identical answers or substantial overlap between ChatGPT-4 and Google (**Figure 1**). Google sources included 40% medical practices, 30% academic, 20% single-surgeon practice, and 10% social media, while ChatGPT-4 used 100% academic sources. Only 3/10 (30%) FAQs with open-ended answers were identical between ChatGPT-4 and Google Web Search (**Figure 2**). Google sources included academic (60%), social media (20%), medical practice (10%), and single-surgeon practice (10%), while 100% of sources for ChatGPT-4 were academic. ChatGPT-4 utilized a higher proportion of academic sources to answer FAQs with open-ended ( $p=0.025$ ) and numeric ( $p=0.001$ ) answers.

**DISCUSSION AND CONCLUSION:** In a replicated patient search for information concerning TSA, ChatGPT-4 provided clinically relevant questions and answers. ChatGPT-4 derived information from academic sources in 100% of cases, while the sources used by Google were heterogeneous. Pending further training and refinement, ChatGPT-4 may be utilized as an adjunct for shoulder surgeons and ancillary staff to decrease the burden of patient inquiries. Currently, ChatGPT-4 should remain used by patients as an adjunct to obtaining information and depend on medical professionals for confirmatory answers.

Figure 1. Top ten internet numeric-based FAQs for TSA and comparison of answers between Google and ChatGPT.

Google Numerical FAQs	Google Answer	ChatGPT Answer
1. What is the average age for shoulder replacement?	60 to 70 years old (Medical Practice)	Over 50 years old (Academic)
2. How long do you have to sleep in a recliner after shoulder surgery?	4 to 6 weeks (Single Surgeon Practice)	4 to 6 weeks (Academic)
3. How long does it take to recover from total shoulder replacement?	At least 6 months (Academic)	Within 3 to 6 months (Academic)
4. How many hours is a total shoulder replacement?	About 3 hours (Academic)	Up to 3 hours (Academic)
5. How long do you wear a sling after total shoulder replacement surgery?	4 to 6 weeks postoperatively (Single Surgeon Practice)	4 to 6 weeks (Academic)
6. What percentage of shoulder replacements are successful?	90% (Medical Practice)	80% to 90% (Academic)
7. How long is the hospital stay for shoulder replacement surgery?	1 to 2 nights (Academic)	1 to 3 days (Academic)
8. How big is the incision for total shoulder replacement?	8 inches (Social Media)	4 to 6 inches (Academic)
9. What is the most painful day after surgery?	Day 2 and 3 (Medical Practice)	First few days after surgery (Academic)
10. How useful is shoulder surgery on a scale from 1 to 10?	8 out of 10 (Medical Practice)	Highly useful (Academic)

Highlighted green cells indicated responses considered identical; yellow cells indicated responses with substantial overlap; red cells indicate responses with no overlap. Parentheses indicate website source category.

Figure 2. Top ten internet FAQs for TSA per Google and ChatGPT.

Google	ChatGPT
1. What is total shoulder arthroplasty? (Technical Details)	1. What is total shoulder arthroplasty (TSA)? (Technical Details)
2. Is total shoulder arthroplasty the same as shoulder replacement? (Technical Details)	2. How is total shoulder arthroplasty performed? (Technical Details)
3. How long does it take to recover from total shoulder replacement? (Timeline of Recovery)	3. Who is a candidate for total shoulder arthroplasty? (Indications/Management)
4. How painful is total shoulder replacement? (Pain)	4. What are the risks and complications associated with total shoulder arthroplasty? (Risks/Complications)
5. Is shoulder arthroplasty a major surgery? (Evaluation of Surgery)	5. How long does it take to recover from total shoulder arthroplasty? (Timeline of Recovery)
6. What is the risk or downside of a shoulder replacement? (Risks/Complications)	6. What is the difference between total shoulder arthroplasty and reverse total shoulder arthroplasty? (Technical Details)
7. What is the average age for shoulder replacement? (Indications/Management)	7. How successful is total shoulder arthroplasty in terms of pain relief and improved function? (Evaluation of Surgery)
8. How long do you have to sleep in a recliner after shoulder surgery? (Restrictions)	8. What is the typical lifespan of a total shoulder arthroplasty implant? (Evaluation of Surgery)
9. Are there permanent restrictions after shoulder replacement? (Restrictions)	9. Are there any alternatives to total shoulder arthroplasty? (Indications/Management)
10. Do you still have a rotator cuff after shoulder replacement? (Technical Details)	10. What types of physical therapy or rehabilitation are recommended after total shoulder arthroplasty? (Indications/Management)

Highlighted green cells indicated responses considered identical; white parentheses indicate the question/topic.