Google Gemini vs ChatGPT: ChatGPT Provides Greater Response Accuracy for Meniscus Tear Frequently Asked Questions

Anthony Farhat¹, Oliver Davis Forst, Abdullah Ghali, Phillip N. Williams²

¹Department of Student Affairs, Baylor College of Medicine, ²Baylor Medicine Orthopedics & Sports Medicine

INTRODUCTION: Artificial intelligence (AI) is revolutionizing orthopedics with the rise of Large Language Models (LLM) such as ChatGPT emerging as an increasingly prominent source of medical information for patients. As patient use of LLM continues to grow, it becomes imperative to assess their medical accuracy to prevent misinformation and prepare clinicians for its implementation into practice. Google's new LLM, Gemini, serves as a competitor to ChatGPT and has recently gained traction. The purpose of this study was to evaluate Google Gemini's reliability against ChatGPT in providing accurate information in response to the most frequently asked questions (FAQs) concerning a prevalent orthopedic injury, meniscal tears. We hypothesize that due to its novelty, Google Gemini will be outperformed by ChatGPT.

METHODS:

Two board-certified orthopedic sports medicine surgeons selected the ten most FAQs concerning meniscus tears that they encountered in clinic from a selection of thirty-one FAQs from sources such as the Hospital for Special Surgery. The finalized ten questions were then posed to ChatGPT and Google Gemini on March 14th, 2024. The answers were recorded and analyzed using the ChatGPT Response Rating System, with a score of 1 being an "excellent response requiring no clarification," a score of 2 being "satisfactory requiring minimal clarification," a score of 3 being "satisfactory requiring substantial clarification" and a score of 4 being "unsatisfactory requiring substantial clarification." The mean score for each chatbot was calculated, and a t-test was performed to compare the two sets of results.

RESULTS: As illustrated in Table 1, Google Gemini earned an average score of 3.1, while ChatGPT had a statistically significant more accurate average rating of 1.8 (p < 0.01). Notable misinformation from Google Gemini and ChatGPT are provided in Table 2.

DISCUSSION AND CONCLUSION:

Overall, our study suggests that ChatGPT outperforms Google Gemini in providing accurate answers to the most FAQs concerning meniscus tears, supporting our hypothesis. Despite the current limitations of ChatGTP and Google Gemini, these free and widely available resources will likely continue to be utilized by patients. As patients continue to turn to AI to investigate their orthopedic injuries, it is important for clinicians to be cognizant of the accuracy of the most frequently used AI platforms to be better equipped to address misleading information among an increasingly educated patient population. Table 1: Summative Performance of Google Gemini and ChatGPT

Question	Gemini Rating	ChatGPT Rating
Do you need surgery for a torn meniscus?	3	1
What is a meniscectomy?	3	1
How long is recovery from a meniscus repair?	2	2
What is the recovery time after meniscectomy?	4	1
How successful is meniscus surgery?	4	3
What are the non-surgical treatments for a torn meniscus?	3	3
What are the surgical treatments for a torn meniscus?	4	4
What is a meniscus tear?	3	1
What are the signs and symptoms of a meniscus tear?	2	1
Can a meniscus tear heal on its own?	3	1
Average Score	3.1	1.8

Google Gemini	ChatGPT	
Physical therapy can independently heal a meniscus tear	Meniscus repair increases the risk of developing osteoarthritis	
Platelet-rich plasma injections can independently heal a meniscus tear	Hyaluronic acid injection is a viable treatment option for meniscus tear healing	
Age is a major prognostic factor for meniscus surgery success	Meniscus transplant and meniscus allograft replacement are different surgeries	
Failing to mention retear or reoperation as a risk of meniscus repair	Failing to mention retear or reoperation as a risk of meniscus repair	