

Comparative Utility Analysis of Chordoma Search Information between ChatGPT vs. Google Web

Shankar S Thiru¹, Addisu Mesfin²

¹Georgetown University School of Medicine, Washington, DC, ²Medstar Orthopaedic Institute, Washington, DC
sst61@georgetown.edu

Disclosures: Shankar S Thiru (N), Addisu Mesfin (4-Axiomed; 6-Medtronic, Stryker, Nuvasive; 9-CSRS, JRGOS, LSRS, NASS, SRS)

INTRODUCTION: There has been a recent surge in the use of artificial intelligence across multiple sectors due to the introduction of ChatGPT in 2022. Consequently, one affected discipline has been the health sphere. ChatGPT is an AI-based large language model (LLM) that is trained through extensive text datasets, enabling it to generate conversational responses to text input (1). Currently, it has started to be used in many facets of the health sphere, such as healthcare education, research, and practice. A Yale study reported that ChatGPT was able to pass the USMLE exams in concordance with the knowledge of a third-year medical student (2). In addition, Barash et al. reported the utility of ChatGPT in assisting with radiologic image selection for clinical cases (3). To our knowledge, the utility of ChatGPT is largely unknown in spine medicine. Moreover, whether ChatGPT is a useful and reliable resource for patients to get information about their spine pathologies remains unanswered. Chordoma is a rare neoplasm derived from notochord remnants with a propensity to metastasize and recur despite conventional, multimodal treatment (4). Its rarity and associated patient hardship lent for an appropriate spine pathology to gauge the utility of ChatGPT in comparison to Google. This study aims to compare the utility of ChatGPT to Google web searching in obtaining information about the spine pathology, chordoma. It is hypothesized that ChatGPT will provide a broader range of relevant questions, as well as more reliable sources for information, due to its unprecedented access to information and integrated, adaptive learning.

METHODS: A web search analysis was performed. With a clean-installed Google Chrome browser, a Google web search of the term was done with the following search term: “chordoma.” The first 10 FAQs were then recorded along with the associated website to each question. The following statements were inputted into ChatGPT: (1) “Perform a google search with the search term ‘chordoma’ and record the 10 FAQs related to the search term” (2) “[INSERT FAQ]. This is my question. Make sure to cite results using [[number](URL)] notation after the reference.” Textual output to the 10 most common FAQs were collected from Google and ChatGPT. Answers and questions produced from the two search modalities were then categorized according to Rothwell question and source classification. Finally, questions that included a numerical response were compiled and inputted into Google and ChatGPT, and answers were recorded.

RESULTS SECTION: There were 3 of 10 questions that were similar amongst FAQs provided by a Google web search and search of ChatGPT for the search term “chordoma.” The most abundant Rothwell question classification from the Google web search, 4 of 10 questions (40%), was “technical details.” The categorization of remaining questions under the Google web search are as follows: timeline of recovery (20%), indications/management (20%), and risk/complications (20%). Pertaining to the search of ChatGPT, the most abundant Rothwell classification was “specific activities,” 3 out of 10 questions (30%). Remaining distribution of categories was technical details (20%), indications/management (20%), risk/complications (20%), and timeline of recovery (10%). In regard to the Google web search, 3 of the 10 questions asked were associated with a response from a commercial website. This contrasts with 1 out of 10 for ChatGPT. In addition, ChatGPT predominantly utilized government sources (70%), the most frequent one being PubMed, for its answers to questions about chordoma. Google’s most abundant source type was academic (50%). All 3 numerical questions (100%) had varied answers between Google and ChatGPT.

DISCUSSION: The study’s results primarily found that the output from a Google web search and a search of ChatGPT substantially differed in questions and answers regarding chordoma. A noticeable distinction in the answers provided by ChatGPT and Google was the length of text. Answers provided by ChatGPT were considerably longer. Important to note is that while having more words, not all text in ChatGPT’s answers directly pertained to the question stem input. There also was a stark contrast in distribution of questions provided by ChatGPT and Google. Of note, ChatGPT provided questions that were classified as “specific activities” that were completely absent in the population of questions provided by Google (30% vs 0%). This is important because one of the “specific activities” questions provided by ChatGPT was: “Are there any support groups or resources available for chordoma patients?” A randomized control study exhibited the value of support groups in increasing the efficacy of heart surgery for positive postoperative outcomes (5). Thus, this additional inclusion of certain “specific activities” questions by ChatGPT suggests that ChatGPT may supply information not readily provided by Google, that could alter recovery and rehabilitation from certain health diseases. This study was not without limitations. First, a small sample size of questions and answers were used, limiting the generalizability of the study. However, due to the novelty of the topic and endeavor, the results still characterize the preliminary utility of ChatGPT as a health information search tool in the field of spine medicine. Another limitation is that ChatGPT’s access to the internet stopped at the year of 2021, while Google has been kept up to date to the present. This is a possible explanation for the difference in questions and answers between the two search tools. Finally, there is some variability in ChatGPT’s replication of answers to a question input because of its AI processing. However, this study found that though the language may vary slightly, the overarching semantics and source of information largely was consistent across multiple iterations of ChatGPT responses to a question. Overall, this data suggests there is some difference amongst chordoma information provided by ChatGPT vs. Google. A comparison of FAQs by a Google web search and search of ChatGPT lent to a heterogeneous sample of questions and responses to open and discrete questioning.

SIGNIFICANCE/CLINICAL RELEVANCE: (1-2 sentences): ChatGPT should remain as a trending resource for patients to obtain information on spine pathologies, such as chordoma, with the caveat that further corroboration and research is needed to verify ChatGPT’s ability to provide credible information that aligns with the goals of the patient and physician alike.

REFERENCES: (1) Sallam M, Giansanti D. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* 2023;11(6):887. (2) Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR medical education* 2023;9:e45312. (3) Barash Y, Klang E, Konen E, Sorin V. ChatGPT-4 Assistance in Optimizing Emergency Department Radiology Referrals and Imaging Selection. *Journal of the American College of Radiology : JACR* 2023. (4) Hoke ATK, Padgett MR, Fabian KP, Nandal A, Gallia GL, Bilusic M, et al. Combinatorial Natural Killer Cell-based Immunotherapy Approaches Selectively Target Chordoma Cancer Stem Cells. *Cancer research communications* 2021;1(3):127-139. (5) Laferton. Enhancing the efficacy of heart surgery by optimizing patients’ preoperative expectations: study protocol of a randomized controlled trial. *Am Heart J* 2013;165(1):1.