ChatGPT for Neurosurgical education

- Transforming neurosurgical practice with large language models: comparative performance of ChatGPT-omni and Gemini in complex case management
- Transforming Neurosurgical Practice with Large Language Models: Comparative Performance of ChatGPT-Omni and Gemini in Complex Case Management
- Accuracy and quality of ChatGPT-4o and Google Gemini performance on image-based neurosurgery board questions
- Performance evaluation of ChatGPT-4.0 and Gemini on image-based neurosurgery board practice questions: A comparative analysis
- Assessing AI Simplification of Medical Texts: Readability and Content Fidelity
- The performance of ChatGPT versus neurosurgery residents in neurosurgical board examination-like questions: a systematic review and meta-analysis
- Examining the Readability of AtlasGPT, the Premiere Resource for Neurosurgical Education
- Large Language Models in Neurosurgery

Arfaie et al. delve into ChatGPT's potential as a learning tool in neurosurgery while contextualizing its abilities for passing medical licensing exams and neurosurgery written boards. Additionally, possibilities for creating personalized case presentations and study material are discussed alongside ChatGPT's capacity to optimize the research workflow and perform a concise literature review. However, such tools need to be used with caution, given the possibility of artificial intelligence hallucinations and other concerns such as user overreliance, and complacency. Overall, this opinion paper raises key points surrounding ChatGPT's role in neurosurgical education [1)]

It is necessary to explore the potential of this program in neurosurgical education and to evaluate its reliability.

While ChatGPT can provide information and answer questions related to neurosurgery, it is crucial to note that it is not a substitute for formal neurosurgical training or medical advice from qualified professionals. Neurosurgery is a highly specialized field that requires extensive education, hands-on experience, and supervision by licensed medical professionals.

A study involved a systematic review of the current literature on the use of artificial intelligence in neurosurgery, with a focus on ChatGPT. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines were followed to ensure a comprehensive and transparent review process. Thirteen studies met the inclusion criteria and were included in the final analysis. The data extracted from the included studies were analyzed and synthesized to provide an overview of the current state of research on the use of ChatGPT in neurosurgery. Results The ChatGPT showed a potential to complement and enhance neurosurgical practice. However, there are risks and limitations associated with its use, including question format limitations, validation challenges, and algorithmic bias. The study highlights the importance of validating machine-generated content for accuracy and addressing ethical concerns associated with AI technologies. The study also identifies potential

benefits of ChatGPT, such as providing personalized treatment plans, supporting surgical planning and navigation, and enhancing large data processing efficiency and accuracy. Conclusion The integration of AI technologies into neurosurgery should be approached with caution and careful consideration of ethical and validation issues. Continued research and development of AI tools in neurosurgery can help us further understand their potential benefits and limitations [2)].

A study explores the potential of ChatGPT, an artificial intelligence model by OpenAI, in enhancing neurosurgical board education. The focus extends beyond technology adoption to its effective utilization, with ChatGPT's proficiency evaluated against practice questions from the Primary Neurosurgery Written Board Exam.

Using the Congress of Neurologic Surgeons (CNS) Self-Assessment Neurosurgery (SANS) Exam Board Review Prep questions, we conducted 3 rounds of analysis with ChatGPT. We developed a novel ChatGPT Neurosurgical Evaluation Matrix (CNEM) to assess the output quality, accuracy, concordance, and clarity of ChatGPT's answers.

ChatGPT achieved spot-on accuracy for 66.7% of prompted questions, 59.4% of unprompted questions, and 63.9% of unprompted questions with a leading phrase. Stratified by topic, accuracy ranged from 50.0% (Vascular) to 78.8% (Neuropathology). In comparison to SANS explanations, ChatGPT output was considered better in 19.1% of questions, equal in 51.6%, and worse in 29.3%. Concordance analysis showed that 95.5% of unprompted ChatGPT outputs and 97.4% of unprompted outputs with a leading phrase were aligned.

The study evaluated the performance of ChatGPT in neurosurgical board education by assessing its accuracy, clarity, and concordance. The findings highlight the potential and challenges of integrating AI technologies like ChatGPT into medical and neurosurgical board education. Further research is needed to refine these tools and optimize their performance for enhanced medical education and patient care [3)].

---

50 questions were included in a written exam, 46 questions were generated by humans (senior staff members) and 4 were generated by ChatGPT. 11 participants took the exam (ChatGPT and 10 residents). Questions were both open-ended and multiple-choice. 8 questions were not submitted to ChatGPT since they contained images or schematic drawings to interpret.

Formulating requests to ChatGPT required an iterative process to precise both questions and answers. Chat GPT scored among the lowest ranks (9/11) among all the participants). There was no difference in response rate for residents between human-generated vs AI-generated questions that could have been attributed to less clarity of the question. ChatGPT answered correctly to all its self-generated questions.

AI is a promising and powerful tool for medical education and specific medical purposes, which need to be further determined. To request AI to generate logical and sound questions, that request must be formulated as precise as possible, framing the content, the type of question, and its correct answers [4)].

In a study, ChatGPT's success in the Turkish Neurosurgical Society Proficiency Board Exams (TNSPBE) is compared to the actual candidates who took the exam, along with identifying the types of questions it answered incorrectly, assessing the quality of its responses, and evaluating its performance based on the difficulty level of the questions. Scores of all 260 candidates were recalculated according to the exams they took and included questions in those exams for ranking purposes of this study. The average score of the candidates for a total of 523 questions is 62.02 ± 0.61 compared to ChatGPT, which was 78.77. They concluded that in addition to ChatGPT's higher response rate, there was also a correlation with the increase in clarity regardless of the difficulty level of the questions with Clarity 1.5, 2.0, 2.5, and 3.0. In the participants, however, there is no such increase in parallel with the increase in clarity [5)].

A study aimed to show the reliability of ChatGPT by asking various questions about the chat engine, how it can contribute to neurosurgery education by preparing case reports or questions, and its contributions when writing academic articles. The results of the study showed that while ChatGPT provided intriguing and interesting responses, it should not be considered a dependable source of information. The absence of citations for scientific queries raises doubts about the credibility of the answers provided. Therefore, it is not advisable to solely rely on ChatGPT as an educational resource. With further updates and more specific prompts, it may be possible to improve its accuracy. In conclusion, while ChatGPT has potential as an educational tool, its reliability needs to be further evaluated and improved before it can be widely adopted in neurosurgical education [6)]

Hegde et al. attempt to generate a short case report of an extraventricular neurocytoma (EVN). When posed with a set of questions on this uncommon topic, ChatGPT did well with relevant but not very specific answers. Validation of the information is an issue like with most machine learning solutions. This case report is an attempt to summarise EVN and its management. Q: Summarise extraventricular neurocytoma. ChatGPT: EVN is a rare and benign tumor that typically occurs in the brain. It often causes no symptoms, but as it grows, it can cause headaches, seizures, and changes in vision or balance. Surgery is usually curative, but there is a lack of studies to know the best treatment options. The prognosis is generally good with high rates of long-term survival and low rates of recurrence [7)].

Furthermore, while the use of ChatGPT in writing scientific articles and other scientific outputs may have potential benefits, important ethical concerns must also be addressed. Consequently, Cascella et al. investigated the feasibility of ChatGPT in clinical and research scenarios: (1) support of the clinical practice, (2) scientific production, (3) misuse in medicine and research, and (4) reasoning about public health topics. Results indicated that it is important to recognize and promote education on the appropriate use and potential pitfalls of AI-based LLMs in medicine [8)].

One potential application of Artificial Intelligence in neurosurgical education is through the development of chatbots or virtual assistants that can provide personalized guidance and support to medical students and residents. These virtual assistants can answer questions, provide feedback, and offer real-time assistance in surgical simulations, helping learners to refine their skills and improve

their knowledge.

Another application of AI in neurosurgical education is through the use of virtual reality (VR) and augmented reality (AR) technologies. These immersive technologies can simulate surgical procedures and allow learners to practice surgical techniques in a safe and controlled environment. By providing a realistic and engaging experience, VR and AR can help to enhance learning and improve the retention of information.

Additionally, AI can be used in neurosurgical education to assist with surgical planning and decision-making. Machine learning algorithms can analyze patient data and help to predict surgical outcomes, allowing neurosurgeons to make more informed decisions and develop more effective treatment plans.

Overall, the use of AI in neurosurgical education has the potential to enhance learning, improve surgical skills, and ultimately lead to better patient outcomes.

GPT-4

GPT-4, an updated language model with additional training parameters, has exhibited exceptional performance on standardized exams. A study examines GPT-4's competence on neurosurgical board-style questions, comparing its performance with medical students and residents, to explore its potential in medical education and clinical decision-making.

GPT-4's performance was examined on 643 Congress of Neurological Surgeons Self-Assessment Neurosurgery Exam (SANS) board-style questions from various neurosurgery subspecialties. Of these, 477 were text-based and 166 contained images. GPT-4 refused to answer 52 questions that contained no text. The remaining 591 questions were inputted into GPT-4, and its performance was evaluated based on first-time responses. Raw scores were analyzed across subspecialties and question types, and then compared to previous findings on Chat Generative pre-trained transformer performance against SANS users, medical students, and neurosurgery residents.

GPT-4 attempted 91.9% of Congress of Neurological Surgeons SANS questions and achieved 76.6% accuracy. The model's accuracy increased to 79.0% for text-only questions. GPT-4 outperformed Chat Generative pre-trained transformer ($P < 0.001$) and scored highest in pain/peripheral nerve (84%) and lowest in spine (73%) categories. It exceeded the performance of medical students (26.3%), neurosurgery residents (61.5%), and the national average of SANS users (69.3%) across all categories.

GPT-4 significantly outperformed medical students, neurosurgery residents, and the national average of SANS users. The mode's accuracy suggests potential applications in educational settings and clinical decision-making, enhancing provider efficiency, and improving patient care [9].

1)
Arfaie S, Sadegh Mashayekhi M, Mofatteh M, Ma C, Ruan R, MacLean MA, Far R, Saini J, Harmsen IE, Duda T, Gomez A, Rebchuk AD, Pingbei Wang A, Rasiah N, Guo E, Fazlollahi AM, Rose Swan E, Amin P, Mohammed S, Atkinson JD, Del Maestro RF, Girgis F, Kumar A, Das S. ChatGPT and neurosurgical education: A crossroads of innovation and opportunity. J Clin Neurosci. 2024 Nov;129:110815. doi: 10.1016/j.jocn.2024.110815. Epub 2024 Sep 4. PMID: 39236407.
2)

Bečulić H, Begagić E, Skomorac R, Mašović A, Selimović E, Pojskić M. ChatGPT's contributions to the evolution of neurosurgical practice and education: a systematic review of benefits, concerns and limitations. Med Glas (Zenica). 2024 Feb 1;21(1). doi: 10.17392/1661-23. Epub ahead of print. PMID: 37950660.
3)

Mannam SS, Subtirelu R, Chauhan D, Ahmad HS, Matache IM, Bryan K, Chitta SVK, Bathula SC, Turlip R, Wathen C, Ghenbot Y, Ajmera S, Blue R, Chen HI, Ali ZS, Malhotra N, Srinivasan V, Ozturk AK, Yoon JW. Large Language Model-Based Neurosurgical Evaluation Matrix: A Novel Scoring Criteria to Assess the Efficacy of ChatGPT as an Educational Tool for Neurosurgery Board Preparation. World Neurosurg. 2023 Dec;180:e765-e773. doi: 10.1016/j.wneu.2023.10.043. Epub 2023 Oct 14. PMID: 37839567.
4)

Bartoli A, May AT, Al-Awadhi A, Schaller K. Probing artificial intelligence in neurosurgical training: ChatGPT takes a neurosurgical residents written exam. Brain Spine. 2023 Nov 29;4:102715. doi: 10.1016/j.bas.2023.102715. PMID: 38163001; PMCID: PMC10753430.
5)

Sahin MC, Sozer A, Kuzucu P, Turkmen T, Sahin MB, Sozer E, Tufek OY, Nernekli K, Emmez H, Celtikci E. Beyond human in neurosurgical exams: ChatGPT's success in the Turkish neurosurgical society proficiency board exams. Comput Biol Med. 2023 Dec 10;169:107807. doi: 10.1016/j.compbiomed.2023.107807. Epub ahead of print. PMID: 38091727.
6)

Sevgi UT, Erol G, Doğruel Y, Sönmez OF, Tubbs RS, Güngor A. The role of an open artificial intelligence platform in modern neurosurgical education: a preliminary study. Neurosurg Rev. 2023 Apr 14;46(1):86. doi: 10.1007/s10143-023-01998-2. PMID: 37059815.
7)

Hegde A, Srinivasan S, Menon G. Extraventricular Neurocytoma of the Posterior Fossa: A Case Report Written by ChatGPT. Cureus. 2023 Mar 6;15(3):e35850. doi: 10.7759/cureus.35850. PMID: 37033498; PMCID: PMC10076908.
8)

Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the Feasibility of ChatGPT in Healthcare: An Analysis of Multiple Clinical and Research Scenarios. J Med Syst. 2023 Mar 4;47(1):33. doi: 10.1007/s10916-023-01925-4. PMID: 36869927; PMCID: PMC9985086.
9)

Guerra GA, Hofmann H, Sobhani S, Hofmann G, Gomez D, Soroudi D, Hopkins BS, Dallas J, Pangal DJ, Cheok S, Nguyen VN, Mack WJ, Zada G. GPT-4 Artificial Intelligence Model Outperforms ChatGPT, Medical Students, and Neurosurgery Residents on Neurosurgery Written Board-Like Questions. World Neurosurg. 2023 Nov;179:e160-e165. doi: 10.1016/j.wneu.2023.08.042. Epub 2023 Aug 18. PMID: 37597659.