

Accuracy Analysis of the American Society of Anesthesiologists Classification Using ChatGPT

Authors : Jae Ni Jang, Young Uk Kim

Abstract : Background: Chat Generative Pre-training Transformer-3 (ChatGPT; San Francisco, California, Open Artificial Intelligence) is an artificial intelligence chatbot based on a large language model designed to generate human-like text. As the usage of ChatGPT is increasing among less knowledgeable patients, medical students, and anesthesia and pain medicine residents or trainees, we aimed to evaluate the accuracy of ChatGPT-3 responses to questions about the American Society of Anesthesiologists (ASA) classification based on patients' underlying diseases and assess the quality of the generated responses. Methods: A total of 47 questions were submitted to ChatGPT using textual prompts. The questions were designed for ChatGPT-3 to provide answers regarding ASA classification in response to common underlying diseases frequently observed in adult patients. In addition, we created 18 questions regarding the ASA classification for pediatric patients and pregnant women. The accuracy of ChatGPT's responses was evaluated by cross-referencing with Miller's Anesthesia, Morgan & Mikhail's Clinical Anesthesiology, and the American Society of Anesthesiologists' ASA Physical Status Classification System (2020). Results: Out of the 47 questions pertaining to adults, ChatGPT -3 provided correct answers for only 23, resulting in an accuracy rate of 48.9%. Furthermore, the responses provided by ChatGPT-3 regarding children and pregnant women were mostly inaccurate, as indicated by a 28% accuracy rate (5 out of 18). Conclusions: ChatGPT provided correct responses to questions relevant to the daily clinical routine of anesthesiologists in approximately half of the cases, while the remaining responses contained errors. Therefore, caution is advised when using ChatGPT to retrieve anesthesia-related information. Although ChatGPT may not yet be suitable for clinical settings, we anticipate significant improvements in ChatGPT and other large language models in the near future. Regular assessments of ChatGPT's ASA classification accuracy are essential due to the evolving nature of ChatGPT as an artificial intelligence entity. This is especially important because ChatGPT has a clinically unacceptable rate of error and hallucination, particularly in pediatric patients and pregnant women. The methodology established in this study may be used to continue evaluating ChatGPT.

Keywords : American Society of Anesthesiologists, artificial intelligence, Chat Generative Pre-training Transformer-3, ChatGPT

Conference Title : ICAIT 2024 : International Conference on Artificial Intelligence and Technology

Conference Location : Athens, Greece

Conference Dates : October 17-18, 2024