ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope

Summary on evaluation of LLM

The capacity of large language models (LLMs) like ChatGPT to assist medical education has been evaluated in Paper 1. The study assessed the performance of ChatGPT on the United States Medical Licensing Exam (USMLE) without any specialized training or reinforcement. ChatGPT demonstrated a high level of concordance and insight in its explanations, suggesting that LLMs may have the potential to assist with medical education and clinical decision-making. However, Paper 2 raises concerns about the potential misuse of ChatGPT as a tool for academic misconduct in online exams. While advanced proctoring techniques and AI-text output detectors may be effective in addressing this issue, they are not likely to be foolproof solutions. Moreover, the accuracy and integrity of generated text produced by ChatGPT in scientific abstracts were evaluated in Paper 3. ChatGPT writes convincing scientific abstracts with completely generated data, though it can be detected using an AI output detector and skeptical human reviewers. It is crucial for journals and medical conferences to adapt policy and practice to maintain rigorous scientific standards by including AI output detectors in the editorial process and clear disclosure if these technologies are used. Finally, Paper 4 suggests that beyond the research community, ChatGPT has gained significant attention in the medical domain for simplifying radiology reports, with potential to improve patient-centered care in radiology and other medical domains.