ChatGPT Displays Remarkable Accuracy in Clinical Decision Making

A recent study, conducted by researchers at Mass General Brigham, has revealed that ChatGPT achieved an overall clinical decision-making accuracy of approximately 72%.

This included various tasks from coming up with potential diagnoses to reaching final diagnoses and making decisions about care management.

The large-language model (LLM) AI chatbot demonstrated consistent performance in both primary care and emergency settings across all medical specialties.

The study provides a comprehensive evaluation of decision support using ChatGPT, spanning from the initial interactions with a patient to the complete care scenario. This assessment encompasses tasks ranging from generating a differential diagnosis to testing, diagnosis, and management.

The study involved pasting 36 standardised clinical vignettes, in which portions were progressively fed into ChatGPT.

Initially, the tool was requested to generate a range of potential differential diagnoses based on the patient's initial information, including age, gender, symptoms, and whether the case was an emergency. Subsequently, ChatGPT received additional data and was requested to formulate management decisions as well as provide a conclusive diagnosis.

To evaluate ChatGPT's performance, the team employed a structured blinded approach, comparing its accuracy on differential diagnosis, diagnostic testing, final diagnosis, and management decisions. Regarding assessment, points were assigned for correct responses and linear regressions were used to analyse the connection between ChatGPT's performance and the vignette's demographic details.

ChatGPT's overall accuracy stood at approximately 72%. Notably, it was best in making a final diagnosis, boasting a success rate of 77%. It was lowest-performing in making differential diagnoses, achieving a rate of only 60%. In terms of clinical management decisions, which included figuring out the most appropriate medications post-accurate diagnosis, ChatGPT displayed a 68% accuracy rate.

Corresponding author Marc Succi, MD, said, "ChatGPT struggled with differential diagnosis, which is the meat and potatoes of medicine when a physician has to figure out what to do".

"That is important because it tells us where physicians are truly experts and adding the most value—in the early stages of patient care with little presenting information, when a list of possible diagnoses is needed".

Although tools like ChatGPT can be considered for integration into clinical practice, further benchmark research and regulatory guidance are essential prerequisites.

Published on : Wed, 23 Aug 2023