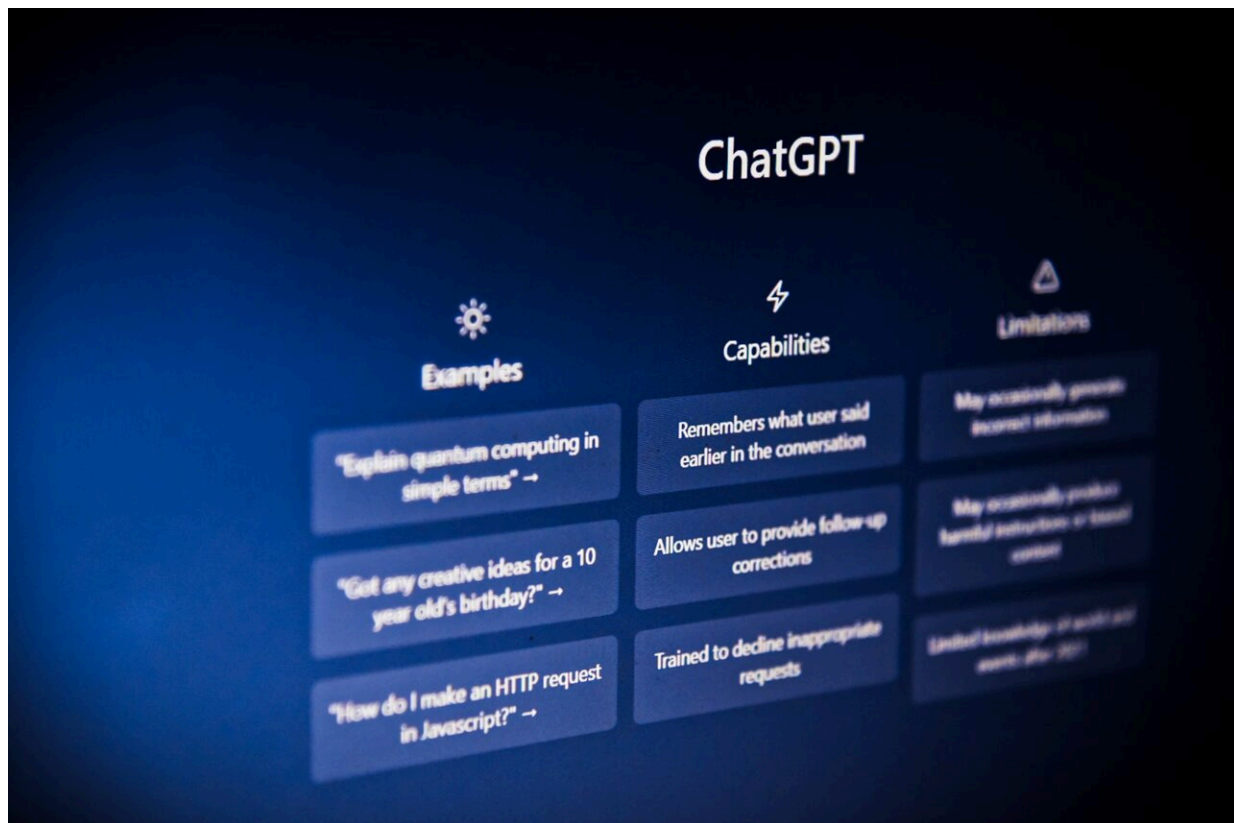# ChatGPT found to have very low success rate in diagnosing pediatric case studies

January 4 2024, by Bob Yirka

Credit: Unsplash/CC0 Public Domain

A trio of pediatricians at Cohen Children's Medical Center, in New York, has found ChatGPT's pediatric diagnostic skills to be considerably lacking after asking the LLM to diagnose 100 random case studies. In

their study, [reported](#) in the journal *JAMA Pediatrics*, Joseph Barile, Alex Margolis and Grace Cason tested ChatGPT's diagnostic skills.

Pediatric diagnostics is particularly challenging, the researchers note, because in addition to taking into account all the symptoms found in a particular patient, age must be considered as well. In this new effort, they noted that LLMs have been promoted by some in the [medical community](#) as a promising new diagnostic tool. To determine their efficacy, the researchers assembled 100 random pediatric case studies and asked ChatGPT to diagnose them.

To keep things simple, the researchers used a single approach in querying the LLM for all the [case studies](#). They first pasted in the text from the [case study](#), and then followed up with the prompt "List a [differential diagnosis](#) and a final diagnosis."

A differential diagnosis is a methodology used to suggest a preliminary diagnosis (or several of them) using a patient's history and physical exams. The final diagnosis, as its name suggests, is the believed cause of the symptoms. Answers given by the LLM were scored by two fellow colleagues who were not otherwise involved in the study—there were three possible scores, "correct," "incorrect" and "did not fully capture diagnosis."

The research team found that ChatGPT produced correct scores just 17 times—of those, 11 were clinically related to the correct [diagnosis](#) but were still wrong.

The researchers note the obvious: ChatGPT is clearly not yet ready to be used as a [diagnostic tool](#), but they also suggest that more selective training could improve results. They further suggest that in the meantime, LLMs like ChatGPT may prove useful as an administrative tool, or to assist in writing research articles or for generating instruction

sheets for patient use in aftercare applications.

More information: Joseph Barile et al, Diagnostic Accuracy of a Large Language Model in Pediatric Case Studies, *JAMA Pediatrics* (2024). DOI: 10.1001/jamapediatrics.2023.5750