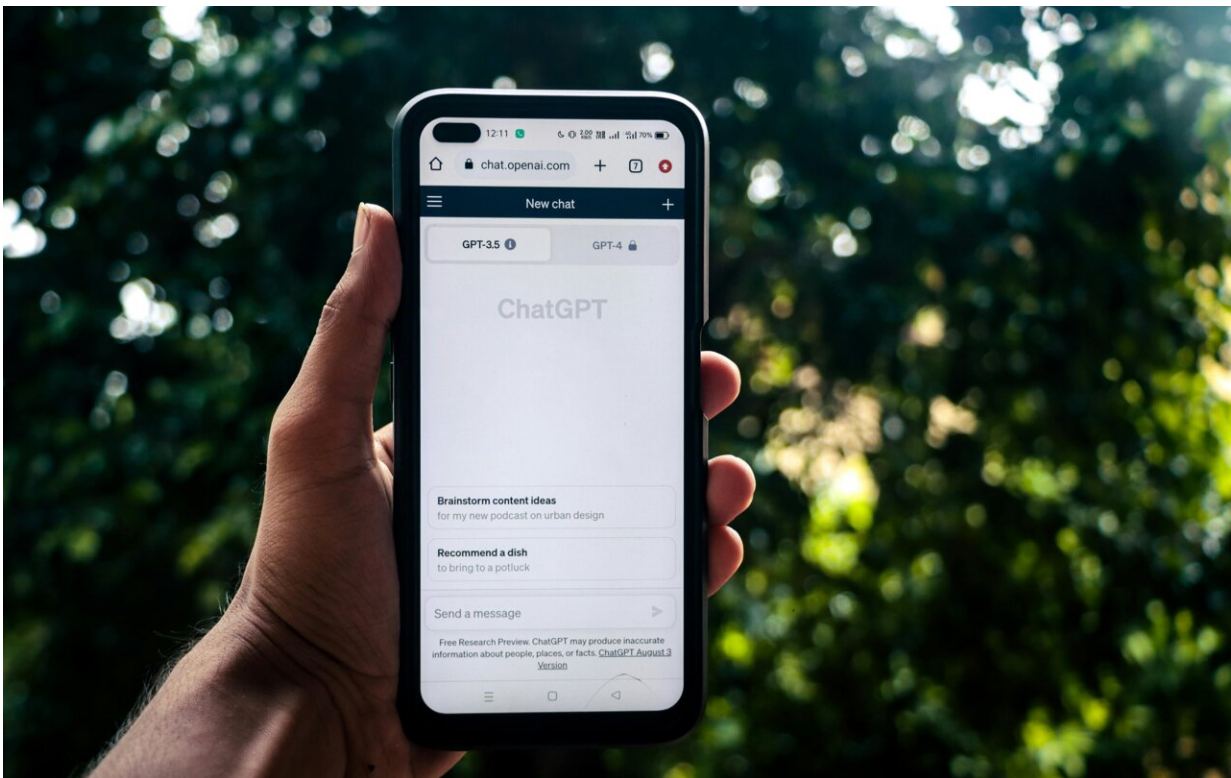# ChatGPT still not very good at diagnosing human ailments

August 1 2024, by Bob Yirka

Credit: Sanket Mishra from Pexels

A team of medical researchers at Western University's Schulich School of Medicine and Dentistry has found that despite being trained on terabytes of data, the LLM ChatGPT is still not good at diagnosing human ailments. In their study, published on the open-access site *PLOS*

ONE, the group trained the popular LLM on 150 case studies and prompted it to provide a diagnosis.

Prior research and anecdotal evidence have shown that LLMs such as ChatGPT can provide impressive results on some prompts, such as to write a love poem for a girlfriend, but it can also return incorrect or bizarre responses. Many in the field have suggested caution when using the results produced by an LLM for important topics like health advice.

For this new study, the team in Canada evaluated how well ChatGPT would diagnose human ailments if given symptoms of real patients as described in actual case studies. They chose 150 case studies from Medscape, an online website created and used by medical professionals for informational and educational purposes, that were accompanied by a known accurate diagnosis. They trained ChatGPT 3.5 with pertinent data, such as patient history, lab results and office exam findings, and then asked it for a diagnosis and/or a treatment plan.

After the LLM returned an answer, the research team graded its results based on how close it came to the correct diagnosis. They also graded it on how well it reported its rationale in reaching its diagnosis, including offering citations—an important part of medical diagnostics. They then averaged the scores received for all the case studies and found that the LLM gave a correct diagnosis just 49% of the time.

The researchers note that while the LLM scored poorly, it did do a good job describing how it reached its diagnosis—a characteristic, the team suggests, that might prove useful for medical students. They also noted that the LLM was reasonably good at ruling out possible ailments. They conclude by suggesting that LLMs are not yet ready for use in diagnostic settings.

More information: Ali Hadi et al, Evaluation of ChatGPT as a

diagnostic tool for medical learners and clinicians, *PLOS ONE* (2024).