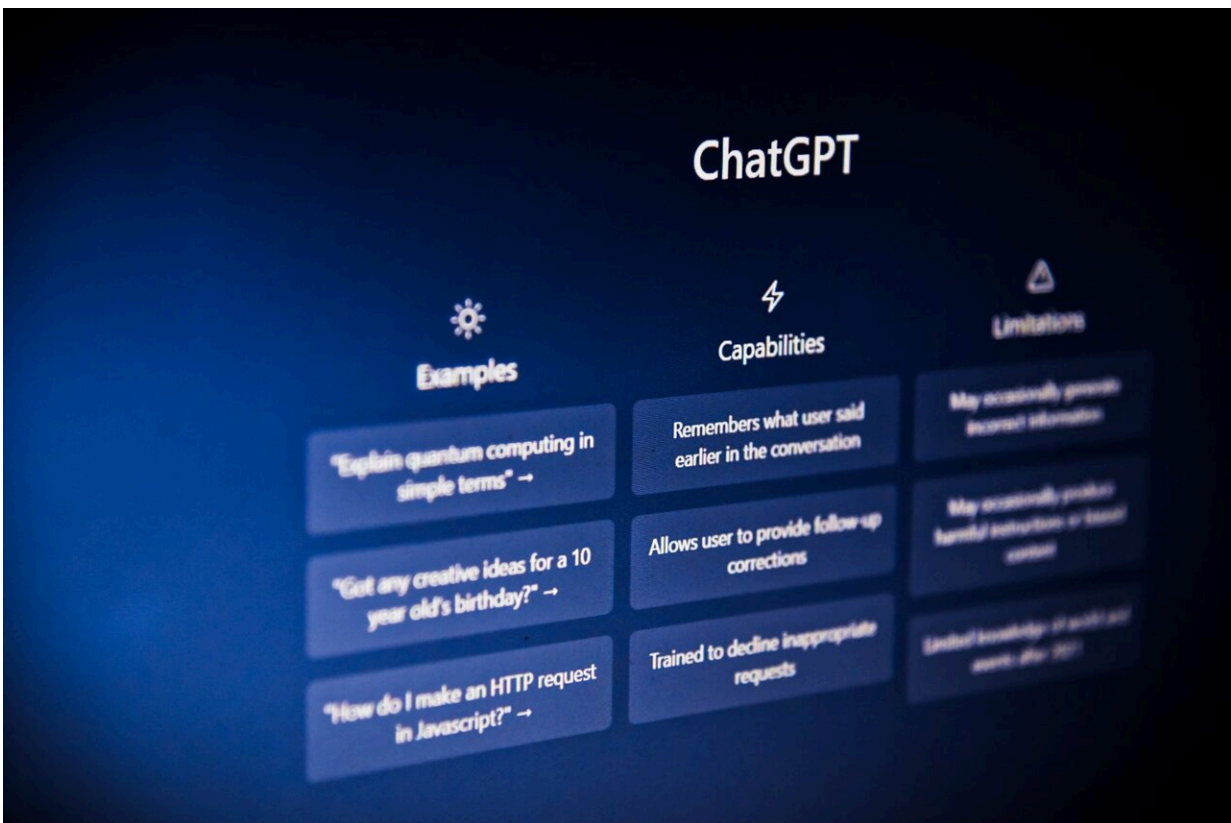


AI language models open a potential Pandora's box of medical research fraud

March 14 2023, by Justin Jackson



Credit: Unsplash/CC0 Public Domain

Medical student and researcher Faisal Elali of the State University of New York Downstate Health Sciences University and medical scribe and researcher Leena Rachid from the New York-Presbyterian/Weill Cornell

Medical Center wanted to see if artificial intelligence could write a fabricated research paper and then investigate how best to detect it.

Artificial intelligence is an increasingly valuable and vital part of scientific research. It is used as a tool to analyze complicated data sets, but it is never used to generate the actual paper for publication. AI-generated [research papers](#), on the other hand, can look convincing even when based on an entirely fabricated study. But exactly how convincing?

In a paper published in the open-access journal *Patterns*, the research duo demonstrated the feasibility of fabricating a research paper using ChatGPT, an AI-based language model. Simply by asking, they were able to have ChatGPT produce a number of well-written, entirely made-up abstracts. A hypothetical fraudster could then submit these fake abstracts to multiple journals seeking publication. If accepted, the same process could be used to write an entire study with false data, nonexistent participants and meaningless results. However, it could appear legitimate, especially if the subject is particularly abstract or not screened by an expert in the specific field.

In [a previous experiment](#) cited in the current paper, humans were given both human-created and AI-generated abstracts to consider. In that experiment, humans incorrectly identified 32% of the AI-generated research abstracts as real and 14% of the human-written abstracts as fake.

The current research team decided to test their ChatGPT fabricated study against three online AI detectors. The texts were overwhelmingly identified as AI-generated, suggesting the adoption of AI detection tools by journals could be a successful diverter of fraudulent applications. However, when they took the same text and ran it through a free, online, AI-powered rephrasing tool first—the consensus unanimously flipped to "likely human," suggesting we need better AI detection tools.

Actual science is hard work, and communicating the details of that work is a crucial aspect of science requiring substantial effort. But any mostly hairless ape can string sensible sounding words together given enough time and coffee—as the writer of this article can firmly attest. Creating a fake study with enough detail to seem credible would take tremendous effort, requiring hours of researching how best to sound believable, and might be too tedious a task for someone interested in malicious mischief. With AI completing the task in minutes, that mischief could become an entirely achievable objective. As the researchers point out in their paper, that mischief could have terrible consequences.

They give an example of a legitimate study that supports the use of drug A over drug B for treating a [medical condition](#). Now, suppose a fabricated study makes the opposite claim and is not detected (as a side note, even if it is detected, clawing back citations and reprints of retracted studies is notoriously difficult). It could impact subsequent meta-analyses and systematic reviews of these studies—studies that guide health care policies, standards of care and clinical recommendations.

Beyond the simple mischief motive, the authors of the paper point to the pressure on [medical professionals](#) to quickly produce a high volume of publications to gain research funding or entry into higher career positions. In part, they point out that the United States Medical Licensing Examination recently switched from a graded exam to a pass/fail model, meaning ambitious students rely more heavily on published research to distinguish them from the pack. This raises the stakes for a trustworthy AI detection system to remove potentially fraudulent medical research that could pollute the publishing environment—or worse still, practitioners who submit fraudulent papers from practicing on patients.

The goal of AI language models has long been to produce texts that are

indistinguishable from human text. That we need AI that can detect when a human is using AI to produce fraudulent work indistinguishable from reality should not come as a surprise. What might be surprising is just that we may need it so soon.

More information: Faisal R. Elali et al, AI-generated research paper fabrication and plagiarism in the scientific community, *Patterns* (2023). DOI: [10.1016/j.patter.2023.100706](https://doi.org/10.1016/j.patter.2023.100706)

© 2023 Science X Network

Citation: AI language models open a potential Pandora's box of medical research fraud (2023, March 14) retrieved 25 April 2024 from <https://medicalxpress.com/news/2023-03-ai-language-potential-pandora-medical.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.