New tool detects fake, AI-produced scientific articles

August 5 2024, by Chris Kocher

Credit: Unsplash/CC0 Public Domain

When ChatGPT and other generative artificial intelligence can produce scientific articles that look real—especially to someone outside that field of research—what's the best way to figure out which ones are fake?

Ahmed Abdeen Hamed, a visiting research fellow at Binghamton University's Thomas J. Watson College of Engineering and Applied Science, has created a [machine-learning algorithm](#) he calls xFakeSci that can detect up to 94% of bogus papers—nearly twice as successfully as more common data-mining techniques.

"My main research is [biomedical informatics](#), but because I work with medical publications, [clinical trials](#), [online resources](#) and mining [social media](#), I'm always concerned about the authenticity of the knowledge somebody is propagating," said Hamed, who is part of George J. Klir Professor of Systems Science Luis M. Rocha's Complex Adaptive Systems and Computational Intelligence Lab.

"Biomedical articles in particular were hit badly during the global pandemic because some people were publicizing false research."

In a new paper [published](#) in the journal *Scientific Reports*, Hamed and collaborator Xindong Wu, a professor at Hefei University of Technology in China, created 50 fake articles for each of three popular medical topics—Alzheimer's, cancer and depression—and compared them to the same number of real articles on the same topics.

Hamed said when he asked ChatGPT for the AI-generated papers, "I tried to use the exact same keywords that I used to extract the literature from the [National Institutes of Health's] PubMed database, so we would have a common basis of comparison. My intuition was that there must be a pattern exhibited in the fake world versus the actual world, but I had no idea what this pattern was."

After some experimentation, he programmed xFakeSci to analyze two major features associated with how the papers were written. One is the numbers of bigrams, which are two words that frequently appear together, such as "[climate change](#)," "clinical trials" or "biomedical

literature." The second is how those bigrams are linked to other words and concepts in the text.

"The first striking thing was that the number of bigrams were very few in the fake world, but in the real world, the bigrams were much more rich," Hamed said. "Also, in the fake world, despite the fact that there were very few bigrams, they were so connected to everything else."

Hamed and Wu theorize that the writing styles are different because human researchers don't have the same goals as AIs prompted to produce a piece on a given topic.

"Because ChatGPT is still limited in its knowledge, it tries to convince you by using the most significant words," Hamed said. "It is not the job of a scientist to make a convincing argument to you. A real research paper reports honestly about what happened during an experiment and the method used. ChatGPT is about depth on a single point, while real science is about breadth."

Distinguished Professor and Chair of the Department of Systems Science and Industrial Engineering Mohammad T. Khasawneh praised Hamed's research.

"We are very glad that the most recent addition to our robust roster of visiting professors, Dr. Ahmed Abdeen Hamed, is working on such novel ideas," he said. "In an era when 'deepfakes' are now part of the general public conversation, his work is incredibly timely and relevant on many levels. We are excited by the promise of his work and look forward to further collaborations with him."

To further develop xFakeSci, Hamed plans to expand the range of topics to see if the telltale word patterns hold for other research areas, going beyond medicine to include engineering, other scientific topics and the

humanities. He also foresees AIs becoming increasingly sophisticated, so determining what is and isn't real will get increasingly difficult.

"We are always going to be playing catchup if we don't design something comprehensive," he said. "We have a lot of work ahead of us to look for a general pattern or universal algorithm that does not depend on which version of generative AI is used."

Because even though their algorithm catches 94% of AI-generated papers, he added, that means 6 out of 100 fakes are still getting through: "We need to be humble about what we've accomplished. We've done something very important by raising awareness."