About Us

Go back

OPINION   02 NOV 2023

Generative AI makes fraud an existential threat to science

By Jorge Quintanilla

Share f t in ✉

Stop judging researchers by their publications or risk a tsunami of fakes, warns Jorge Quintanilla

The ability of generative artificial intelligence to produce text and images indistinguishable from the work of humans has both gripped the public imagination and caused alarm. In the creative professions, the strain is beginning to show—witness, for example, this year's strikes by Hollywood actors and writers.

Scientists might think that their jobs are safe: after all, our task is to extract information from the natural world, not to create it. Generative AI, the argument runs, might have some niche applications in scientific work, or aid menial tasks such as retouching the draft of a paper, but it does not go to the core of the scientific process.

This view is dangerously complacent.

In the creative industries, generative AI tools such as ChatGPT and Midjourney threaten disruption not because they can produce writing or images that are better than a human might come up with, but because they can produce them at scale. If AI can do a scriptwriter's job a thousand times faster at a thousandth of the cost, a human needs to be a million times better to justify their fee.

So it is not the quality of the output but the scale at which it can be produced that is of concern. This is relevant to science, too, once you realise that there are fraudsters in the midst of the scientific community.

Infinite possibilities

In 2002, Julia Hsu and Lynn Loo, then working at Bell Labs in the United States, noticed two identical plots in separate papers co-authored by Jan Hendrik Schön, one of that legendary lab's most prolific scientists. This observation lifted the lid on one of the most notorious frauds in the history of physics.

Over several years, Schön had been fooling his collaborators, their bosses and the scientific community at large into believing that he had achieved a series of breakthroughs—published in Science or Nature every few weeks—that would revolutionise electronics. While some groups had expressed frustration at how difficult it was to reproduce his results, others suspected he was leaving crucial details out to cover his tracks from competitors on the way to a Nobel prize.

Unfortunately for Schön, he couldn't resist the temptation to produce ever-more apparent discoveries at an ever-faster rate. In the end, he got sloppy, using the same data for more than one paper. That's how he got caught.

Imagine how much further a modern-day fraudster could get, armed with generative AI. They wouldn't need to reuse old data. They could produce infinite amounts of seemingly genuine data, each graph tailored to each new fraudulent paper, non-stop.

Tipping point

Scientific journals are not taking the threat from generative AI lightly. Nature requires authors to detail the use of AI tools in the preparation of a manuscript, and has banned the listing of an AI as a co-author.

This is welcome, but unscrupulous authors will simply not own up. Generative AI will also

make some ways of detecting fraudulent papers—such as requesting full data sets—of little use.

There is an argument that incorrect or made-up science is pathological and gets detected by the normal operation of the scientific process. As evidence, we can point to the Schön affair or the recent flurry of activity to reproduce—or, more to the point, fail to reproduce—claims of room-temperature superconductivity.

But such a belief fails to recognise generative AI's ability to scale fraud. A single, inexperienced individual may use the technology to produce an entire, plausible paper, including data, bibliography and text, in a matter of minutes. And that makes all the difference.

We do not have a predictive, quantitative theory of the scientific process. But it is reasonable to assume that, like any complex, self-correcting process, there is an error threshold beyond which it breaks down.

One feature of better-understood self-correcting systems, from self-replicating biological molecules to error-corrected quantum algorithms, is that they don't cross this threshold gradually, but suffer something akin to a phase transition, like the sudden freezing of water when it drops below zero-degrees celsius.

As the error rate rises, self-correcting systems keep working, but slow down as effort is diverted from useful information processing into repair. But when a certain threshold is reached, failure is sudden: cells cannot replicate their DNA; quantum processors lose coherence; or we cease to have a working scientific process capable of the cumulative advancement of knowledge.

Old problem, new danger

This is the nightmare scenario: generative AI empowers dishonest scientists to forge data and papers with such ease and at such a rate that self-correction mechanisms are overwhelmed, and the scientific process enters a new, dysfunctional state.

Preventing this starts with acknowledging the problem. We should consider that, even though the overwhelming majority of scientists are honest, vast swathes of the scientific literature are potentially fraudulent—and expect this fraction to get bigger really fast. We should also acknowledge the risk that unless this trend is curbed, a critical error threshold may be reached.

Finally, and most importantly, we must tackle the root cause of the problem: an excessive reliance on papers as the end product of the scientific process.

An individual or institution's contribution to science should not be equated with the numbers of papers they have published, the journals they published in or the citations their papers received. It is not even about how well-written or interesting those papers

seem. The real contribution is the discoveries and insights that result from their research.

The totemisation of the scientific paper has left science as vulnerable to generative AI as scriptwriting or graphic art, because a paper describing the results of actual scientific work and one that has been made up may look very much alike.

This problem has been recognised for some time: from the signatories of the San Francisco Declaration on Research Assessment in 2013 to the recommendations by the International Advisory Group to the UK's Future Research Assessment Programme in 2023, there is a growing effort to judge science not on its form, but on its underlying contents.

It is difficult to come up with alternative ways to quantify and assess scientific research. Generative AI is making the challenge much more urgent.

Jorge Quintanilla is a reader in condensed matter physics at the University of Kent, Canterbury

Related news

OPINION   25 MAY 2023

Using AI in peer review

OPINION   17 MAY 2023

Combat paper mills with slow science, not warring AIs

OPINION   15 SEP 2022

Automating citations risks encouraging sloppy science

POLITICAL SCIENCE BLOG   28 OCT 2021

AI in research could be a rocket-booster or a treadmill

＊

Visit *Research Professional website

Contact Us

Advertise

Privacy Policy

Terms & Conditions

Copyright

Manage cookie preferences

Jobs

LinkedIn

Research Professional News

Research Fortnight

Research Europe

Research Africa

Cookie list