

# A new 'AI scientist' can write science papers without any human input—here's why that's a problem

August 21 2024, by Karin Verspoor

---



Credit: Wes Cockx & Google DeepMind / Better Images of AI, [CC BY](#)

Scientific discovery is one of the most sophisticated human activities. First, scientists must understand the existing knowledge and identify a significant gap. Next, they must formulate a research question and design and conduct an experiment in pursuit of an answer. Then, they must analyze and interpret the results of the experiment, which may raise yet another research question.

Can a process this complex be automated? Last week, [Sakana AI Labs announced](#) the creation of an "AI scientist"—an artificial intelligence system they claim can make scientific discoveries in the area of [machine learning](#) in a fully automated way.

Using generative large language models (LLMs) like those behind ChatGPT and other AI chatbots, the system can brainstorm, select a promising idea, code new algorithms, plot results, and write a paper summarizing the experiment and its findings, complete with references. Sakana claims the AI tool can undertake the complete lifecycle of a scientific experiment at a cost of just US\$15 per paper—less than the cost of a scientist's lunch.

These are some big claims. Do they stack up? And even if they do, would an army of AI scientists churning out [research papers](#) with inhuman speed really be good news for [science](#)?

## **How a computer can 'do science'**

A lot of science is done in the open, and almost all [scientific knowledge](#) has been written down somewhere (or we wouldn't have a way to "know" it). Millions of scientific papers are freely available online in repositories such as [arXiv](#) and [PubMed](#).

LLMs trained with this data capture the language of science and its patterns. It is therefore perhaps not at all surprising that a generative LLM can produce something that looks like a good scientific paper—it has ingested many examples that it can copy.

What is less clear is whether an AI system can produce an interesting scientific paper. Crucially, good science requires novelty.

## But is it interesting?

Scientists don't want to be told about things that are already known. Rather, they want to learn new things, especially new things that are significantly different from what is already known. This requires judgment about the scope and value of a contribution.

The Sakana system tries to address interestingness in two ways. First, it "scores" new paper ideas for similarity to existing research (indexed in the [Semantic Scholar](#) repository). Anything too similar is discarded.

Second, Sakana's system introduces a "peer review" step—using another LLM to judge the quality and novelty of the generated paper. Here again, there are plenty of examples of peer review online on sites such as [openreview.net](#) that can guide how to critique a paper. LLMs have ingested these, too.

## AI may be a poor judge of AI output

Feedback is mixed on Sakana AI's output. Some have described it as producing "[endless scientific slop](#)."

Even the system's own review of its outputs judges the papers weak at best. This is likely to improve as the technology evolves, but the question of whether automated scientific papers are valuable remains.

The ability of LLMs to judge the quality of research is also an open question. My own work (soon to be published in [Research Synthesis Methods](#)) shows LLMs are not great at judging the risk of bias in medical research studies, though this too may improve over time.

Sakana's system automates discoveries in computational research, which

is much easier than in other types of science that require physical experiments. Sakana's experiments are done with code, which is also structured text that LLMs can be trained to generate.

## **AI tools to support scientists, not replace them**

AI researchers have been developing systems to support science for decades. Given the huge volumes of published research, even finding publications relevant to a specific scientific question can be challenging.

Specialized [search tools](#) make use of AI to help scientists find and synthesize existing work. These include the above-mentioned Semantic Scholar, but also newer systems such as [Elicit](#), [Research Rabbit](#), [scite](#) and [Consensus](#).

Text mining tools such as [PubTator](#) dig deeper into papers to identify key points of focus, such as specific genetic mutations and diseases, and their established relationships. This is especially useful for curating and organizing scientific information.

Machine learning has also been used to support the synthesis and analysis of medical evidence, in tools such as [Robot Reviewer](#). Summaries that compare and contrast claims in papers from [Scholarcy](#) help to perform literature reviews.

All these tools aim to help scientists do their jobs more effectively, not to replace them.

## **AI research may exacerbate existing problems**

While Sakana AI [states](#) it doesn't see the role of human scientists diminishing, the company's vision of "a fully AI-driven scientific

ecosystem" would have major implications for science.

One concern is that, if AI-generated papers flood the scientific literature, future AI systems may be trained on AI output and undergo [model collapse](#). This means they may become increasingly ineffectual at innovating.

However, the implications for science go well beyond impacts on AI science systems themselves.

There are already bad actors in science, including "paper mills" churning out [fake papers](#). This problem will only [get worse](#) when a scientific paper can be produced with US\$15 and a vague initial prompt.

The need to check for errors in a mountain of automatically generated research could rapidly overwhelm the capacity of actual scientists. The peer review system is arguably already broken, and dumping more research of questionable quality into the system won't fix it.

Science is fundamentally based on trust. Scientists emphasize the integrity of the scientific process so we can be confident our understanding of the world (and now, the world's machines) is valid and improving.

A scientific ecosystem where AI systems are key players raises fundamental questions about the meaning and value of this process, and what level of trust we should have in AI scientists. Is this the kind of scientific ecosystem we want?

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

## Provided by The Conversation

Citation: A new 'AI scientist' can write science papers without any human input—here's why that's a problem (2024, August 21) retrieved 3 September 2024 from

<https://techxplore.com/news/2024-08-ai-scientist-science-papers-human.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.