

# Hypotheses devised by AI could find ‘blind spots’ in research - Nature - 17-11-2023

Artificial intelligence is asking questions that humans hope to answer.

By [Matthew Hutson](#)

In early October, as the Nobel Foundation announced the recipients of this year’s Nobel prizes, a group of researchers, including a previous laureate, met in Stockholm to discuss how artificial intelligence (AI) might have an increasingly creative role in the scientific process. The workshop, led in part by Hiroaki Kitano, a biologist and chief executive of Sony AI in Tokyo, considered creating prizes for AIs and AI–human collaborations that produce world-class science. Two years earlier, Kitano proposed the Nobel Turing Challenge<sup>1</sup>: the creation of highly autonomous systems (‘AI scientists’) with the potential to make Nobel-worthy discoveries by 2050.

It’s easy to imagine that AI could perform some of the necessary steps in scientific discovery. Researchers already use it to search the literature, automate data collection, run statistical analyses and even draft parts of papers. Generating hypotheses — a task that typically requires a creative spark to ask interesting and important questions — poses a more complex challenge. For Sendhil Mullainathan, an economist at the University of Chicago Booth School of Business in Illinois, “it’s probably been the single most exhilarating kind of research I’ve ever done in my life”.

## Network effects

AI systems capable of generating hypotheses go back more than four decades. In the 1980s, Don Swanson, an information scientist at the University of Chicago, pioneered literature-based discovery — a text-mining exercise that aimed to sift ‘undiscovered public knowledge’ from the scientific literature. If some research papers say that A causes B, and others that B causes C, for example, one might hypothesize that A causes C. Swanson created software called Arrowsmith that searched collections of published papers for such indirect connections and proposed, for instance, that fish oil, which reduces blood viscosity, might treat Raynaud’s syndrome, in which blood vessels narrow in response to cold<sup>2</sup>. Subsequent experiments proved the hypothesis correct.

Literature-based discovery and other computational techniques can organize existing findings into ‘knowledge graphs’, networks of nodes representing, say, molecules and properties. AI can analyse these networks and propose undiscovered links between molecule nodes and property nodes. This process powers much of modern drug discovery, as well as the task of assigning functions to genes. A review article published in *Nature*<sup>3</sup> earlier this year explores other ways in which AI has generated hypotheses, such as proposing simple formulae that can organize noisy data points and predicting how proteins will fold up. Researchers have automated hypothesis generation in particle physics, materials science, biology, chemistry and other fields.

One approach is to use AI to help scientists brainstorm. This is a task that large language models — AI systems trained on large amounts of text to produce new text — are well suited for, says Yolanda Gil, a computer scientist at the University of Southern California in Los Angeles who has worked on AI scientists. Language models can produce inaccurate information and present it as real, but this ‘hallucination’ isn’t necessarily bad, Mullainathan says. It signifies, he says, “‘here’s a kind of thing that looks true’. That’s exactly what a hypothesis is.”

Blind spots are where AI might prove most useful. James Evans, a sociologist at the University of Chicago, has pushed AI to make ‘alien’ hypotheses — those that a human would be unlikely to make. In a paper published earlier this year in *Nature Human Behaviour*<sup>4</sup>, he and his colleague Jamshid Sourati built knowledge graphs containing not just materials and properties, but also researchers. Evans and Sourati’s algorithm traversed these networks, looking for hidden shortcuts between materials and properties. The aim was to maximize the plausibility of AI-devised hypotheses being true while minimizing the chances that researchers would hit on them naturally. For instance, if scientists who are studying a particular drug are only distantly connected to those studying a disease that it might cure, then the drug’s potential would ordinarily take much longer to discover.

When Evans and Sourati fed data published up to 2001 to their AI, they found that about 30% of its predictions about drug repurposing and the electrical properties of materials had been uncovered by researchers, roughly six to ten years later. The system can be tuned to make predictions that are more likely to be correct but also less of a leap, on the basis of concurrent findings and collaborations, Evans says. But “if we’re predicting what people are going to do next year, that just feels like a scoop machine”, he adds. He’s more interested in how the technology can take science in entirely new directions.

## Keep it simple

Scientific hypotheses lie on a spectrum, from the concrete and specific (‘this protein will fold up in this way’) to the abstract and general (‘gravity accelerates all objects that have mass’). Until now, AI has produced more of the former. There’s another spectrum of hypotheses, partially aligned with the first, which ranges from the uninterpretable (these thousand factors lead to this result) to the clear (a simple formula or sentence). Evans argues that if a machine makes useful predictions about individual cases — “if you get all of these particular chemicals together, boom, you get this very strange effect” — but can’t explain why those cases work, that’s a technological feat rather than science. Mullainathan makes a similar point. In some fields, the underlying principles, such as the mechanics of protein folding, are understood and scientists just want AI to solve the practical problem of running complex computations that determine how bits of proteins will move around. But in fields in which the fundamentals remain hidden, such as medicine and social science, scientists want AI to identify rules that can be applied to fresh situations, Mullainathan says.

In a paper presented in September<sup>5</sup> at the Economics of Artificial Intelligence Conference in Toronto, Canada, Mullainathan and Jens Ludwig, an economist at the University of Chicago, described a method for AI and humans to collaboratively generate broad, clear hypotheses. In a proof of concept, they sought hypotheses related to characteristics of defendants’ faces that might influence a judge’s decision to free or detain them before trial. Given mugshots of past defendants, as well the judges’ decisions, an algorithm found that numerous subtle facial features correlated with judges’ decisions. The AI generated new mugshots with those features cranked either up or down, and human participants were asked to describe the general differences between them. Defendants likely to be freed were found to be more “well-groomed” and “heavy-faced”. Mullainathan says the method could be applied to other

complex data sets, such as electrocardiograms, to find markers of an impending heart attack that doctors might not otherwise know to look for. “I love that paper,” Evans says. “That’s an interesting class of hypothesis generation.”

In science, experimentation and hypothesis generation often form an iterative cycle: a researcher asks a question, collects data and adjusts the question or asks a fresh one. Ross King, a computer scientist at Chalmers University of Technology in Gothenburg, Sweden, aims to complete this loop by building robotic systems that can perform experiments using mechanized arms<sup>6</sup>. One system, called Adam, automated experiments on microbe growth. Another, called Eve, tackled drug discovery. In one experiment, Eve helped to reveal the mechanism by which a toothpaste ingredient called triclosan can be used to fight malaria.

## Robot scientists

King is now developing Genesis, a robotic system that experiments with yeast. Genesis will formulate and test hypotheses related to the biology of yeast by growing actual yeast cells in 10,000 bioreactors at a time, adjusting factors such as environmental conditions or making genome edits, and measuring characteristics such as gene expression. Conceivably, the hypotheses could involve many subtle factors, but King says they tend to involve a single gene or protein whose effects mirror those in human cells, which would make the discoveries potentially applicable in drug development. King, who is on the organizing committee of the Nobel Turing Challenge, says that these “robot scientists” have the potential to be more consistent, unbiased, cheap, efficient and transparent than humans.

Researchers see several hurdles to and opportunities for progress. AI systems that generate hypotheses often rely on machine learning, which usually requires a lot of data. Making more papers and data sets openly available would help, but scientists also need to build AI that doesn’t just operate by matching patterns but can also reason about the physical world, says Rose Yu, a computer scientist at the University of California, San Diego. Gil agrees that AI systems should not be driven only by data — they should also be guided by known laws. “That’s a very powerful way to include scientific knowledge into AI systems,” she says.

As data gathering becomes more automated, Evans predicts that automating hypothesis generation will become increasingly important. Giant telescopes and robotic labs collect more measurements than humans can handle. “We naturally have to scale up intelligent, adaptive questions”, he says, “if we don’t want to waste that capacity.”

Hypotheses devised by AI could find ‘blind spots’ in research

**Artificial intelligence is asking questions that humans hope to answer.**

<https://www.nature.com/articles/d41586-023-03596-0>

## References

1. 1.

Kitano, H. *npj Syst. Biol. Appl.***7**, 29 (2021).

[Article](#) [PubMed](#) [Google Scholar](#)

2. 2.

Swanson, D. R. *Perspect. Biol. Med.***30**, 7–18 (1986).

[Article](#) [PubMed](#) [Google Scholar](#)

3. 3.

Wang, H. *et al. Nature***620**, 47–60 (2023).

[Article](#) [PubMed](#) [Google Scholar](#)

4. 4.

Sourati, J. & Evans, J. A. *Nature Hum. Behav.***7**, 1682–1696 (2023).

[Article](#) [PubMed](#) [Google Scholar](#)

5. 5.

Ludwig, J. & Mullainathan, S. Working Paper 31017 (National Bureau of Economic Research, 2023).

6. 6.

King, R., Peter, O. & Courtney, P. in *Artificial Intelligence in Science* 129–139 (OECD Publishing, 2023).