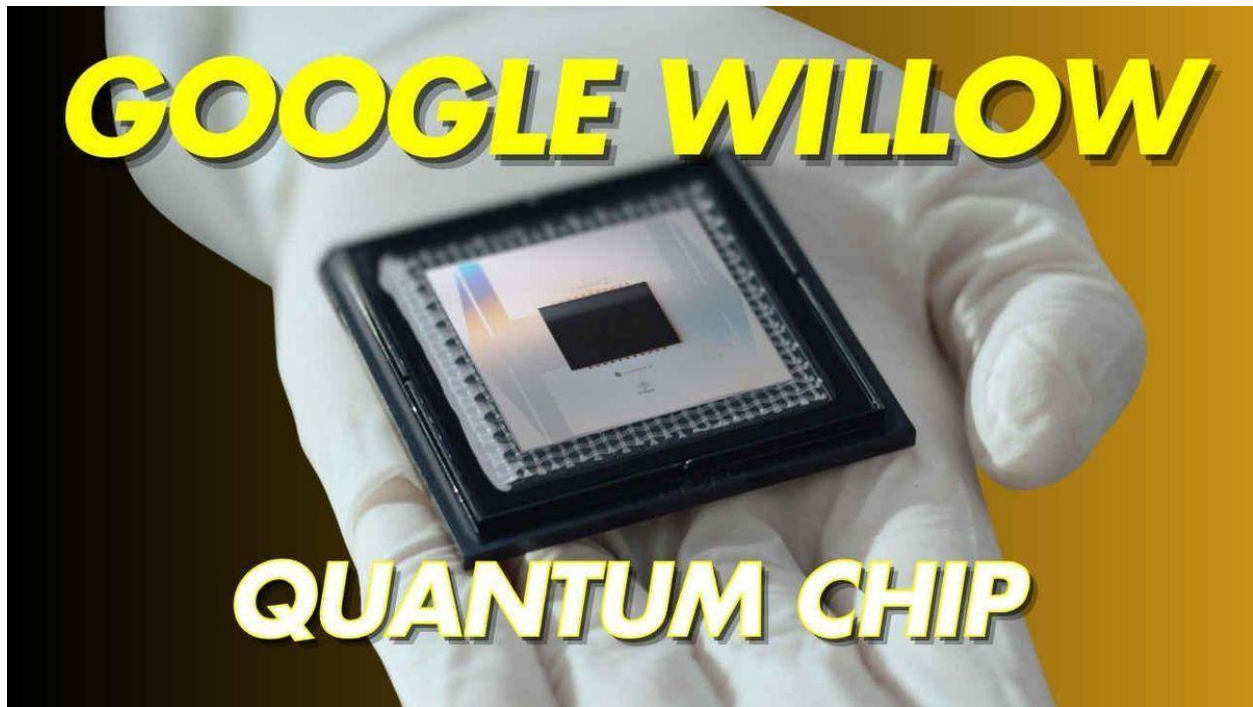




Google's Quantum Chip Can Do in 5 Minutes What Would Take Other Computers 10 Septillion Years

Google's quantum computing division unveiled a new chip, dubbed Willow, that the tech giant says makes it infinitely faster and better than existing supercomputers.

The chip addresses two major problems with quantum computers as we know them today. First, Google claims Willow can reduce quantum computing errors "exponentially." Quantum computers, as Microsoft's Azure Quantum [explains](#), are more prone to errors than classical computers because they exist in fragile states.

And researchers believe that quantum computers need more qubits to be more powerful—but more qubits also means more errors. Tackling this error issue has long been one of the quantum computing sector's largest challenges—one Google may be able to tackle.

"The more qubits we use in Willow, the more we *reduce* errors, and the more quantum the system becomes," [writes](#) Google Quantum AI lead Hartmut Neven (a qubit is a basic unit of quantum information, and the building block for a quantum computer).

The Willow chip. (Credit: Google.)

"We tested ever-larger arrays of physical qubits, scaling up from a grid of 3x3 encoded qubits, to a grid of 5x5, to a grid of 7x7—and each time, using our latest advances in quantum error correction, we were able to cut the error rate in half. In other words, we achieved an exponential reduction in the error rate," Neven continued.

Google has been [working](#) on quantum chips for the past decade. Its latest Willow chip is so fast that it can complete a computation in less than five minutes that would take a leading supercomputer of today like the [Frontier](#) supercomputer in Tennessee 10 septillion years, or 10,000,000,000,000,000,000,000,000 years, which is technically older than the universe itself, according to the company.

"It lends credence to the notion that quantum computation occurs in many parallel universes, in line with the idea that we live in a multiverse," Neven said, citing David Deutsch's prediction detailed in his 1997 book *The Fabric of Reality*.

In 2019, Google previously did a quantum computation in 200 seconds that would have taken the fastest supercomputer at that time 10,000 years. But IBM [dismissed](#) its claims, arguing that a classical computer could do the same calculation with greater "fidelity" in 2.5 days and argued that "quantum supremacy" had yet to be achieved.

While quantum computers are arguably still in their early stages, some researchers have already started to show how such machines might be able to disrupt our lives in the future—or potentially make them better. Chinese researchers have begun to [crack](#) military-grade with one machine, partly affirming a theory that quantum computers could eventually make present-day encryption methods obsolete.

But quantum computers could also help us better understand the universe, and could solve difficult computational problems. As the industry develops, some quantum firms are also looking to [establish manufacturing in North America](#) to ensure American and Canadian national security.

Google's 'Willow' chip aimed at leap in quantum computing



Quantum computers like the one here at the Leibniz Supercomputing Center can calculate much faster than other computers, but they also make many mistakes.© Sven Hoppe/dpa/picture alliance

In some scientific fields, such as data encryption, drug and materials research or machine learning and artificial intelligence, conventional or "classical" computers have reached their limits.

The calculations that researchers need in these fields are taking an infinitely long time to complete, or have become simply impossible.

That's why researchers around the world are working to tap into the enormous potential of quantum computers — they could solve mathematical problems much faster than the fastest conventional computers.

What is special about quantum computers?

The potential of current computing technology has been completely exhausted.

Computers work with bits that know only two states: 1 for "on" and 0 for "off".

[Quantum bits](#), also known as qubits, on the other hand, can assume an infinite number of states.

As a result, quantum computers [can calculate much faster](#), but are also more prone to errors.

"Disturbances from the outside, such as thermal radiation or noise from the electronics, or from the inside, such as material defects" can cause calculation errors, according to Stefan Filipp, a professor in quantum computing at the Technical University of Munich (TUM).

"Although these are now very small — one thousandth to one ten-thousandth of every operation goes wrong — they are not yet small enough [for quantum computers] to run sufficiently long algorithms, " says Filipp.

What has Google achieved with its 'Willow' chip?

The development of quantum computers is [still in its infancy](#). Researchers are currently working with 100 qubits of computing power. Soon there will be 10,000 qubits, and 10,000,000 qubits are also conceivable in the future.

But the systems that have been developed so far are too small and make too many mistakes to deliver any real value. And their error rate tends to increase with additional computing power.

Without a reliable way to correct such errors, quantum computers [will be useless in the real-world](#).

Markus Müller, Professor of Theoretical Quantum Technology at RWTH Aachen University, said the Google team succeeded for the first time in achieving quantum error correction with error rates below a recognized threshold.

That's led to their developing Willow. German computer scientist Hartmut Neven wrote in the journal Nature that with Willow, Google was paving the way for

quantum computing to become common. Neven is the founder and director of Google's Quantum Artificial Intelligence Laboratory.

Will quantum computers arrive soon?

The short answer is "no". While researchers in the field say Google has taken a significant step, Google concedes that Willow's error rate is still too high for quantum computing to be useful in an applied setting.

Willow consists of 105 superconducting qubits. That is less than 10% of the 1,457 physical qubits they say they need to achieve satisfactory error rates.

"In order to ultimately be able to calculate virtually error-free and thus reliably, the error correction rate must increase faster than the error rate," said Michael Hartmann, a professor of theoretical physics at the Friedrich Alexander University in Erlangen-Nürnberg.

"With the current quality of qubits, you will need 100,000 to one million qubits to be able to perform large, fault-tolerant calculations," said Hartmann. This shows, he said, how far quantum computing still has to travel.

This article was translated from the original piece written in German

Source:

Neven H et al. (2024): Quantum error correction below the surface code threshold. Nature. DOI: 10.1038/s41586-024-08449-y.

Google claims quantum supremacy again

Google has once again claimed the prestigious milestone of quantum supremacy, asserting that its latest quantum computer has outperformed classical counterparts in specific computational tasks. This achievement, however, has sparked a heated debate within the scientific community, with some experts raising skepticism about the validity and implications of Google's claims.

Understanding Quantum Supremacy

Definition and Historical Context

Quantum supremacy is a term used to describe the moment when a quantum computer can perform a calculation that is impossible or infeasible for a classical computer. This milestone is significant as it marks a pivotal point in the field of quantum computing, indicating that quantum machines have reached a new level of capability. Google's previous claim of quantum supremacy in 2019, using its Sycamore processor, stirred the tech industry by asserting that it had completed a task in 200 seconds that would take the world's most powerful supercomputer 10,000 years.

Google's initial claim catalyzed a surge in quantum computing research and investment, pushing other companies to accelerate their quantum efforts. Although their 2019 claim faced criticism and skepticism, it undeniably highlighted the potential of quantum technology and set the stage for further developments in the field.

Criteria for Quantum Supremacy

Determining quantum supremacy involves benchmarks that compare the performance of quantum and classical computers on particular tasks. These tasks are often carefully selected to highlight the strengths of quantum systems, which can sometimes lead to criticism. A major challenge is ensuring these tasks have real-world relevance, as critics argue that some chosen benchmarks do not translate into practical applications.

The criteria for quantum supremacy have been a topic of controversy, with some experts questioning their fairness and applicability. The scientific community continues to debate the best standards for evaluating quantum advancements, emphasizing the need for transparent and reproducible results.

Google's Latest Quantum Breakthrough

The Willow Quantum Computer

Google's new quantum computer, named Willow, represents a significant leap over previous systems. With advanced technical specifications, Willow is designed to handle more complex computations and a higher number of qubits, which are the fundamental units of quantum information. This machine is seen as a testament to Google's commitment to advancing quantum technology.

Willow showcases critical improvements in error rates and coherence times, addressing some of the limitations that have plagued earlier quantum systems. These advancements make Willow a powerful tool for exploring the potential of quantum computing in solving problems that are currently intractable for classical computers.

Performance and Achievements

One of the notable achievements of the Willow quantum computer is its ability to solve a complex optimization problem significantly faster than classical computers. This task, which involves intricate calculations, highlights Willow's potential to surpass traditional computing methods. Google's claim of quantum supremacy relies on these kinds of benchmarks to demonstrate the quantum computer's superiority.

The specific tasks Willow has accomplished have reignited discussions about the benchmarks used to claim quantum supremacy. While Google's achievement is undeniably impressive, it also underscores the ongoing debate about the criteria for determining such breakthroughs in the quantum realm.

Industry Skepticism and Criticism

Reactions from Competitors

The announcement of Google's latest achievement has not gone unnoticed by its competitors in the quantum computing arena. Companies like D-Wave have expressed skepticism about Google's claims. D-Wave, for instance, has been vocal about its own advancements, as seen in their recent [assertion of achieving quantum supremacy](#) before others, a claim that has sparked its own debates.

Competing claims of quantum supremacy, such as those from IBM and D-Wave, highlight the fiercely competitive nature of the industry. Each company employs different approaches and technologies, contributing to a diverse landscape of strategies and innovations.

Scientific Community's Response

The scientific community has responded to Google's announcement with a mix of intrigue and skepticism. Some researchers question the validity of Google's results, pointing out potential issues with reproducibility and the practicality of the tasks used to demonstrate supremacy. For instance, a [recent article](#) challenges the robustness of Google's claims, emphasizing the need for careful scrutiny.

Concerns about the real-world applicability of Google's achievements are prevalent among experts. While the tasks completed by Willow are impressive, critics argue that they do not necessarily translate into practical applications, leaving the door open for further debate on the true impact of quantum supremacy claims.

Implications for the Future of Quantum Computing

Potential Applications

The advancements made by Google's Willow quantum computer could have transformative effects on various industries. Quantum computing has the potential to revolutionize fields like cryptography, where it could break current encryption methods, and drug discovery, where it might accelerate the development of new medications. These potential applications underscore the importance of continued research and development in quantum technologies.

As quantum computing progresses, its impact may extend to areas such as logistics optimization, financial modeling, and artificial intelligence. The ability to solve complex problems more efficiently could lead to significant breakthroughs across these domains, making quantum computing a critical area of focus for future technological advancements.

The Road Ahead

Despite Google's impressive achievements, significant challenges remain in the development of quantum computing. Issues such as error correction, scalability, and the development of practical applications continue to pose hurdles for researchers. Addressing these challenges is crucial for realizing the full potential of quantum technology.

Looking forward, the quest for quantum supremacy will likely involve a combination of technological advancements and collaborative efforts. The future of quantum computing promises exciting possibilities, but it also demands careful consideration of the ethical and practical implications of these powerful technologies.

The Competitive Landscape of Quantum Computing

Key Players and Strategies

The quantum computing industry is characterized by a diverse set of players, each pursuing unique strategies to achieve breakthroughs. Major companies like Google, IBM, and D-Wave are at the forefront of this technological race. Google's approach, focusing on developing more powerful quantum processors, contrasts with D-Wave's emphasis on [annealing-based quantum systems](#).

Each of these companies employs different technologies and methodologies, contributing to a vibrant and competitive landscape. These differences highlight the multifaceted nature of quantum computing and the various paths to achieving quantum supremacy.

Innovations and Collaborations

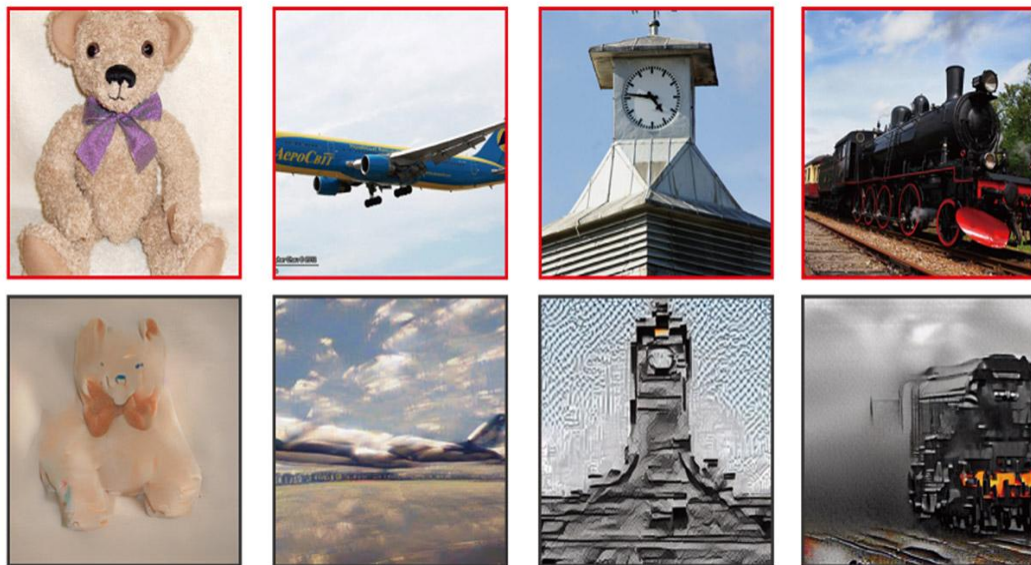
Recent innovations in quantum computing have been driven by both individual company efforts and collaborative initiatives. Companies are increasingly engaging in partnerships with academic institutions and government agencies to accelerate research and development. These collaborations are crucial for pooling resources and expertise, fostering an environment of shared knowledge and rapid progress.

As the field continues to evolve, the role of collaborations will likely become even more significant. By working together, organizations can overcome the challenges inherent in quantum computing and pave the way for groundbreaking advancements. For instance, Google's recent developments have been a topic of discussion, as seen in a [detailed analysis](#) of Willow's impact on the industry.

AI re-creates what people see by reading their brain scans

A new artificial intelligence system can reconstruct images a person saw based on their brain activity

7 MAR 2023



Artificial intelligence re-creations of images based on brain scans (bottom row) match the layout, perspective, and contents of the actual photos seen by study participants (top row). CREATIVE COMMONS

As neuroscientists struggle to demystify how the human brain converts what our eyes see into mental images, artificial intelligence (AI) has been getting better at mimicking that feat. A [recent study](#), scheduled to be presented at an upcoming computer vision conference, demonstrates that AI can read brain scans and re-create largely realistic versions of images a person has seen. As this technology develops, researchers say, it could have numerous applications, from exploring how various animal species perceive the world to perhaps one day recording human dreams and aiding communication in people with paralysis.

Many labs have used AI to read brain scans and re-create images a subject has recently seen, such as [human faces](#) and [photos of landscapes](#). The new study marks the first time an AI algorithm called [Stable Diffusion](#), developed by a German group and publicly

released in 2022, has been used to do this. Stable Diffusion is similar to other text-to-image “generative” AIs such as [DALL-E 2](#) and [Midjourney](#), which produce new images from text prompts after being trained on billions of images associated with text descriptions.

For the new study, a group in Japan added additional training to the standard Stable Diffusion system, linking additional text descriptions about thousands of photos to brain patterns elicited when those photos were observed by participants in brain scan studies.

Unlike previous efforts using AI algorithms to decipher brain scans, which had to be trained on large data sets, Stable Diffusion was able to get more out of less training for each participant by incorporating photo captions into the algorithm. It’s a novel approach that incorporates textual and visual information to “decipher the brain,” says Ariel Goldstein, a cognitive neuroscientist at Princeton University who was not involved with the work.

The AI algorithm makes use of information gathered from different regions of the brain involved in image perception, such as the occipital and temporal lobes, according to Yu Takagi, a systems neuroscientist at Osaka University who worked on the experiment. The system interpreted information from functional magnetic resonance imaging (fMRI) brain scans, which detect changes in blood flow to active regions of the brain. When people look at a photo, the temporal lobes predominantly register information about the contents of the image (people, objects, or scenery), whereas the occipital lobe predominantly registers information about layout and perspective, such as the scale and position of the contents. All of this information is recorded by the fMRI as it captures peaks in brain activity, and these patterns can then be reconverted into an imitation image using AI.

In the new study, the researchers added additional training to the Stable Diffusion algorithm using an online [data set](#) provided by the University of Minnesota, which consisted of brain scans from four participants as they each viewed a set of 10,000 photos. A portion of these brain scans from the same four participants were not used in training and were used to test the AI system later.

ADVERTISEMENT

Each AI-generated image starts out as noise, reminiscent of TV static, and replaces the noise with distinguishable features as the Stable Diffusion algorithm compares a person’s brain activity patterns from viewing a photo with the patterns in its training data set. The system effectively generates an image depicting the contents, layout, and perspective of the photo being viewed. Takagi says the new system was more efficient

than previous ones, required less fine-tuning, and could be trained with a smaller data set.

Brain activity, predominantly in the occipital lobe, provided enough information to re-create the layout and perspective of the photos being viewed, the researchers found. But the algorithm struggled to recapitulate objects, such as a clock tower, from the real photo and instead created abstract figures. One approach to tackling this problem would be to use larger training data sets that could train the algorithm to predict more details, but the fMRI data set was too limited for this, the Japanese team says.

Instead, the researchers circumvented this issue by harnessing keywords from image captions that accompanied the photos in the Minnesota fMRI data set. If, for example, one of the training photos contained a clock tower, the pattern of brain activity from the scan would be associated with that object. This meant that if the same brain pattern was exhibited once more by the study participant during the testing stage, the system would feed the object's keyword into Stable Diffusion's normal text-to-image generator and a clock tower would be incorporated into the re-created image, following the layout and perspective indicated by the brain pattern, resulting in a convincing imitation of the real photo.



Stable Diffusion can re-create photos (left) seen by study participants. Using patterns of brain activity alone, it correctly reproduces the layout and perspective (middle), but with the addition of textual information, it can also correctly re-create the object in the photo (right). CREATIVE COMMONS

Importantly, the stable diffusion algorithm doesn't receive a text prompt directly from the test data—it can only infer that an object is present if the brain pattern matches one seen in the training data. This limits the objects it can re-create to those present in the photos used during training.

Finally, the researchers tested their system on additional brain scans from the same participants when they viewed a separate set of photos, including a toy bear, airplane, clock, and train. By comparing the brain patterns from those images with those produced by the photos in the training data set, the AI system was able to produce convincing imitations of the novel photos. (The team posted a [preprint](#) of its work in December 2022.)

“The accuracy of this new method is impressive,” says Iris Groen, a neuroscientist at the University of Amsterdam who was not involved with the work.

However, the AI system was only tested on brain scans from the same four people who provided the training brain scans, and expanding it to other individuals would require retraining the system on their brain scans. So, it may take a while for this technology to become widely accessible. Nonetheless, Groen argues that “these diffusion models have [an] unprecedented ability to generate realistic images,” and could create new opportunities for cognitive neuroscience research.

Shinji Nishimoto, another systems neuroscientist at Osaka University who worked on the study, hopes that with further refinements the technology could be used to intercept imagined thoughts and dreams, or could allow scientists to understand how differently other animals perceive reality.

Scientists design algorithm that 'reads' people's thoughts from brain scans published October 24, 2022

Using fMRI, scientists decoded what people were hearing and thinking.



Scientists used a computer algorithm and a brain scanning technique called fMRI to "read" peoples' minds. (This is a stock image.) (Image credit: Aja Koska via Getty Images)

Scientists can now "decode" people's thoughts without even touching their heads, [The Scientist reported](#).

Past mind-reading techniques relied on [implanting electrodes](#) deep in peoples' brains. The new method, described in a report posted Sept. 29 to the preprint database bioRxiv, instead relies on a [noninvasive brain scanning technique](#) called functional magnetic resonance imaging (fMRI). fMRI tracks the flow of oxygenated blood through the [brain](#), and because active brain cells need more energy and oxygen, this information provides an indirect measure of brain activity.

By its nature, this scanning method cannot capture real-time brain activity, since the electrical signals released by brain cells move much more quickly than blood moves through the brain. But remarkably, the study authors found that they could still use this

imperfect proxy measure to decode the semantic meaning of people's thoughts, although they couldn't produce word-for-word translations.

"If you had asked any cognitive neuroscientist in the world 20 years ago if this was doable, they would have laughed you out of the room," senior author [Alexander Huth](#), a neuroscientist at the University of Texas at Austin, told The Scientist.

For the new study, which has not yet been peer-reviewed, the team scanned the brains of one woman and two men in their 20s and 30s. Each participant listened to 16 total hours of different podcasts and radio shows over several sessions in the scanner. The team then fed these scans to a computer algorithm that they called a "decoder," which compared patterns in the audio to patterns in the recorded brain activity.

The algorithm could then take an fMRI recording and generate a story based on its content, and that story would match the original plot of the podcast or radio show "pretty well," Huth told The Scientist. In other words, the decoder could infer what story each participant had heard based on their brain activity. That said, the algorithm did make some mistakes, like switching up characters' pronouns and the use of the first and third person. It "knows what's happening pretty accurately, but not who is doing the things," Huth said.

Sign up for the Live Science daily newsletter now

Get the world's most fascinating discoveries delivered straight to your inbox.

Contact me with news and offers from other Future brandsReceive email from us on behalf of our trusted partners or sponsors

By submitting your information you agree to the [Terms & Conditions](#) and [Privacy Policy](#) and are aged 16 or over.

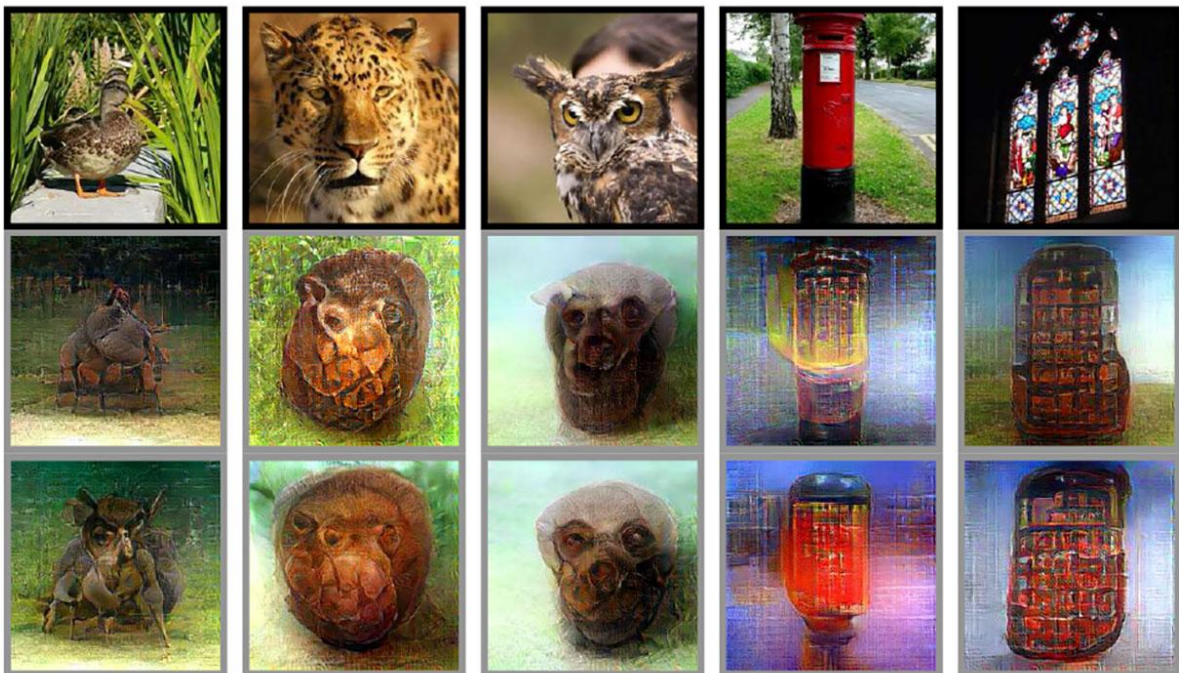
In additional tests, the algorithm could fairly accurately explain the plot of a silent movie that the participants watched in the scanner. It could even retell a story that the participants imagined telling in their heads. In the long term, the research team aims to develop this technology so that it can be used in brain-computer interfaces designed for people who cannot speak or type.

Read more about the new decoder algorithm in [The Scientist](#).

This 'mind-reading' algorithm can decode the pictures in your head

New computer program uses brain activity to draw images of airplanes, leopards, and stained-glass windows

10 JAN 2018



Imagine searching through your digital photos by mentally picturing the person or image you want. Or sketching a new kitchen design without lifting a pen. Or texting a loved one a sunset photo that was never captured on camera. A computer that can read your mind would find many uses in daily life, not to mention for those paralyzed and with no other way to communicate. Now, scientists have created the first algorithm of its kind to interpret—and accurately reproduce—images seen or imagined by another person. It might be decades before the technology is ready for practical use, but researchers are one step closer to building systems that could help us project our inner mind's eye outward.

"I was impressed that it works so well," says Zhongming Liu, a computer scientist at Purdue University in West Lafayette, Indiana, who helped develop an algorithm that

can [somewhat reproduce what moviegoers see when they're watching a film](#). "This is really cool."

Using algorithms to decode mental images isn't new. Since 2011, researchers have recreated movie clips, photos, and even dream imagery by matching brain activity to activity recorded earlier when viewing images. But these methods all have their limits: Some deal only with narrow domains like face shape, and others can't build an image from scratch—instead, they must select from preprogrammed images or categories like "person" or "bird." This new work can generate recognizable images on the fly and even reproduce shapes that are not seen, but imagined.

To figure out what a person is seeing, researchers turned to functional magnetic resonance imaging (fMRI), which measures blood flow to the brain as a proxy for neural activity. They mapped out visual processing areas to a resolution of 2 millimeters as three people looked at more than 1000 images several times each. The goal was to consider just the activity in response to an image—like a leopard—and eventually have a computer paint an image that would produce nearly the same activity.

But instead of showing their subjects painting after painting until the computer got it right, the team built a software stand-in for the brain, a [deep neural network](#) (DNN) with several layers of simple processing elements. "We believe that a deep neural network is good proxy for the brain's hierarchical processing," says Yukiyasu Kamitani, a neuroscientist at Kyoto University in Japan, and the study's senior author. "By using a DNN we can extract information from different levels of the brain's visual system," from simple light contrast up to more meaningful content such as faces.

Using a "decoder," the researchers created representations of the brain's responses to the images, but in the DNN. From then on, they no longer needed the actual fMRI measurements, just the DNN translations.

In guessing what someone was viewing, the translation acts as a template, and the fMRI data are set aside. The system then tries to paint a picture that will trigger the DNN to respond in a way that matches this template. It does so through trial and error until, hopefully, it paints the desired image, whether a leopard, a duck, or a stained-glass window. The system starts with something random—similar to TV static—and slowly refines its painting over the course of 200 rounds. To get closer to the ideal image, the system calculates the difference between the DNN activity and the templated DNN activity. Those calculations cause it to nudge one pixel this way and one pixel that way, until it gets closer to its ideal image.

To make the final product more accurate, researchers included a "deep generator network" (DGN), an algorithm that in this case has been pretrained to generate realistic images based on its input. The DGN refines the paintings to look more naturalistic. Once that was added, a neutral human observer could tell which of two photos an image was meant to recreate [99% of the time](#), the researchers reported in a paper uploaded to the preprint server bioRxiv late last month.

Next, the scientists tried to read the minds of people merely imagining images. This time they scanned the three subjects' brains after asking them to recall images previously displayed, including a fish, an airplane, and simple colored shapes. The method didn't work well for photos, but for the shapes, the generator created a recognizable image 83% of the time.

It's "interesting and careful work," says Nikolaus Kriegeskorte, a computational neuroscientist at Columbia University's Zuckerman Institute. He wonders to what extent the inaccuracies in the computer-generated images are due to limitations in brain activity measurements and to what extent they reflect mistakes in how our brains interpret images. "Higher-resolution fMRI and other brain imaging techniques might further improve the results," he says. With better measurements and continued improvement in algorithms, we might someday communicate through mental pictures.

Scientists are using MRI scans to reveal the physical makeup of our thoughts and feelings November 24, 2019

Who among us hasn't wished we could read someone else's mind, know exactly what they're thinking? Well that's impossible, of course, since our thoughts are, more than anything else, our own. Private, personal, unreachable. Or at least that's what we've always, well, thought.

Advances in neuroscience have shown that, on a physical level, our thoughts are actually a vast network of neurons firing all across our brains. So if that brain activity could be identified and analyzed, could our thoughts be decoded? Could our minds be read? Well, a team of

scientists at Carnegie Mellon University in Pittsburgh has spent more than a decade trying to do just that. We started our reporting on their work 10 years ago, and what they've discovered since, has drawn us back.



In Carnegie Mellon's scanner room, two floors underground, a steady stream of research subjects come to have their brains and thoughts "read" in this MRI machine. It's a type of scanning called functional MRI, fMRI.

That looks at what's happening inside the brain as a person thinks.

Marcel Just: It's like being an astronomer when the first telescope is discovered, or being a biologist when the first microscope is-- is developed. Neuroscientist Marcel Just says this technology has made it possible for the first time to see the physical makeup of our thoughts.

When we first visited Dr. Just's lab ten years ago, he and his team had conducted a study. They put people in the scanner and asked them to think about ten objects, five of them tools like screwdriver and hammer

and five of them dwellings like igloo and castle, while measuring activity levels throughout their brains. The idea was to crunch the data and try to identify distinctive patterns of activity for each object.

Lesley Stahl: You had them think about a screwdriver.

Marcel Just: Uh-huh.

Lesley Stahl: And the computer found the place in the brain where that person was thinking "screwdriver?"

Marcel Just: Screwdriver isn't one place in the brain. It's many places in the brain. When you think of a screwdriver, you think about how you hold it, how you twist it, what it looks like...

Lesley Stahl: And each of those functions are in different places?

Marcel Just: Correct.

He showed us that by dividing the brain into thousands of tiny cubes and analyzing the amount of activity in each one, his team was able to identify unique patterns for each object.

Lesley Stahl: You're reading their mind.

Marcel Just: We're identifying the thought that's occurring. It's--

Lesley Stahl: Whoa.

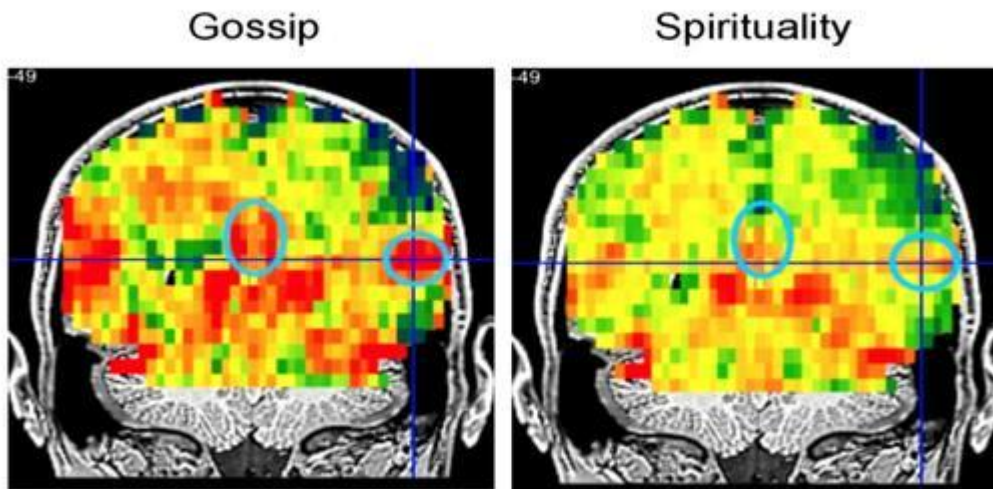
Marcel Just: --incredible, just incredible.



Marcel Just

Incredible, but only the beginning. In the decade since, Professor Just's lab has taken this technique and applied it far beyond hammers and igloos, to increasingly complex thoughts. This is basic science, knowledge for knowledge's sake. Not trying to cure disease, but to understand the fundamental workings of our bodies, and in this case, of our minds.

One of Dr. Just's main questions was whether he could find patterns for abstract ideas, so he did a study asking people to think about forgiveness, gossip, spirituality. Could they be identifiable in the brain the way the screwdriver was?



Remarkably, the answer was yes. This was the activation pattern when people thought about spirituality. And this was gossip.

Lesley Stahl: One of my favorite subjects.

Marcel Just: And you see a slightly different pattern.

One difference between the two was in areas of the brain scientists had already shown become active when we think about other people. Circled in blue. Those areas lit up bright red when subjects thought about gossip; not so much for spirituality.

In another study, Dr. Just tested whether patterns are the same when people think in different languages. They are. And he's asked acting students to conjure up emotions in the scanner to see if feelings have distinctive activation patterns too.

Lesley Stahl: And what did you find?

Marcel Just: Each emotion had its own characteristic values. And you could tell which one was which.

Lesley Stahl: And it's the same in every head.

Marcel Just: Amazingly, it was common across people.

Common across people? Does that mean we could put our colleague, associate producer Jaime Woods, into a scanner for the first time, and Dr. Just's team would be able to identify her emotions?

Lesley Stahl: So she's seeing words.

For nine seconds each, Jaime's job was to think of little scenarios that would conjure up the feelings on the screen. After she came out, a computer program took the brain activation data gathered by the scanner and tried to decode her thoughts.

Lesley Stahl: So what were you thinking about for disgust?

Jaime Woods: I was thinking of someone throwing up on me (LAUGH) at, like, a baseball game.

So could the computer read her brain patterns and tell what she'd been feeling?

COMPUTER VOICE: The program's answer is "I think the emotion is disgust." The experienced emotion was actually disgust. That was correct.

Jaime Woods: Awesome.

Next...

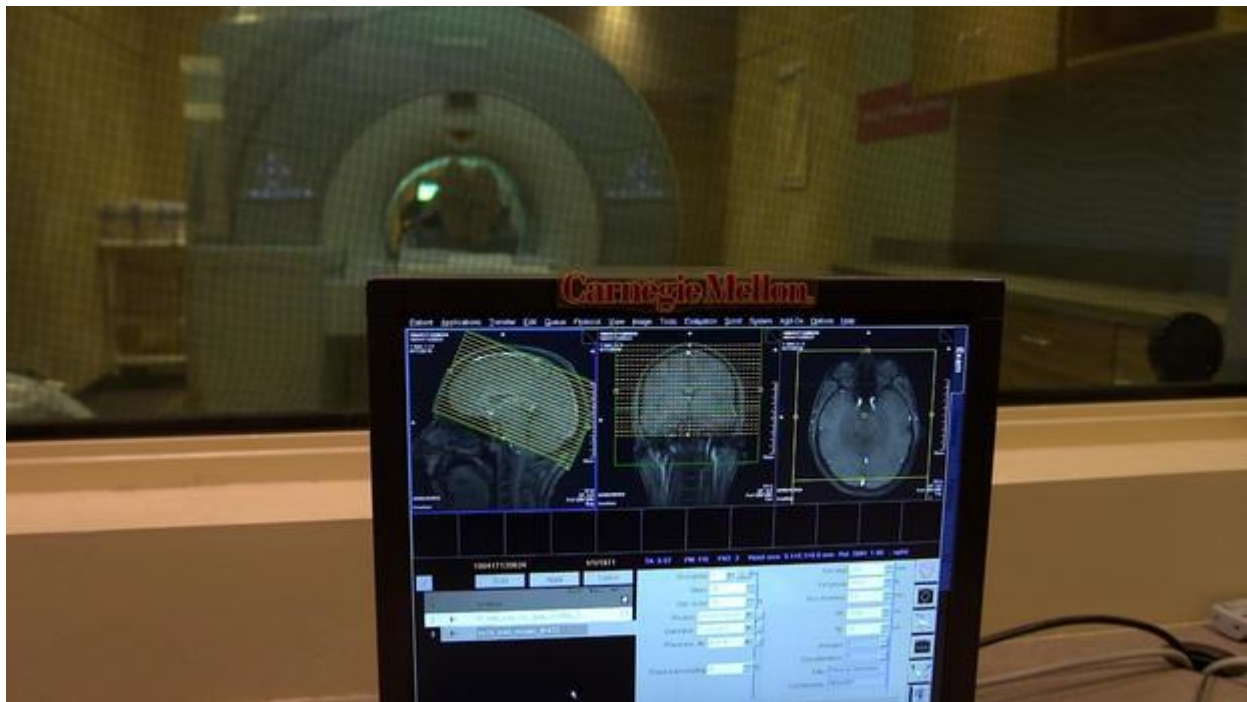
COMPUTER VOICE: "I think the emotion is envy."

Lesley Stahl: What were you thinking for envy?

Jaime Woods: I was just thinking of beautiful models.

The computer program got all of Jaime's emotions right.

Lesley Stahl: It's reading what Jaime's feeling.



Marcel Just: And it's funny isn't it because it's so personal. We all think of our own thoughts as so individual, so intimate, how could anybody else's thoughts be like mine? And they are.

Lesley Stahl: It's feelings, too.

Marcel Just: Yes, feelings. Now, obviously, people think very different thoughts. But, you know-- like, people choose to do different things with their bodies. But they all walk putting one foot in front of the other. Nobody walks sideways. Nobody walks backwards systematically.

There's something about the biological apparatus that makes you act in a certain way with your body. And I don't think we realize the degree to which the biological apparatus that we have in our skulls governs, shapes the way we think.

Professor Just's goal is to one day create a dictionary of brain activation, a key to what all different thoughts look like inside our minds. But he also started wondering whether those definitions might be different in people with disorders like autism.

Prior research had found structural differences in the brains of people with autism, so the question was whether thought patterns might differ too.

Dr. Just's team recruited 17 adults with autism and asked them, as well as 17 control subjects, to think about social interactions like adore, hug, humiliate. Challenging terrain for many with autism. The results were striking, the activation patterns differed enough to tell who had autism and who didn't with 97% accuracy.

Marcel Just: The people with autism thought of these social interactions apart from themselves.

As he showed us in findings for the word "hug," the key differences were in brain regions that activate when we think about ourselves, circled in blue.

Lesley Stahl: Right there.

Marcel Just: There, and there. Those areas light up much more among the controls.

Whereas the autism subjects showed far less activation.

Marcel Just: They thought of it more like a definition of "hug" without self involvement.

Lesley Stahl: And that, you saw, with word after word?

Marcel Just: Yes.



David Brent

David Brent: I just thought, "Wow, this is the coolest thing I've heard in I don't know how long."

David Brent is a psychiatrist at the University of Pittsburgh Medical Center, where he runs a clinic for suicidal adolescents. He happened to attend a talk Marcel Just was giving about his autism findings and immediately wondered about his own patients.

David Brent: So I went up to him afterwards and I said, "Would you be interested in talking about maybe doing a study on suicide?"

Lesley Stahl: You hear about cases of suicide where the person had been depressed, but you also hear of situations where people say, "There was nothing wrong."

David Brent: Suicide is a great mystery, because the person who knows the most about why it happened isn't there to talk with. You try to reconstruct what happened but nobody has a window into people's, you know, interior-- thoughts.

Nobody, that is, except someone with a mind-reading device. Dr. Just and Dr. Brent began planning a pilot study to see if the scanner might reveal what is altered in the thoughts of people contemplating suicide. They reached out to Matt Nock, a Harvard professor who has studied how difficult it is for doctors and emergency rooms to know which patients are safe to send home.

Lesley Stahl: Is this the first time anybody's looked inside the brain to see-- about suicidal thoughts?

Matt Nock: Yes. This is the first study I-- I have ever heard of where someone's looked in the brain of someone who's suicidal, who's actively thinking about death or suicide--

Dan Toski: You don't see life as something that's going to be fixed. The only way to get out of it is to kill yourself.

Dan Toski, a former patient at Dr. Brent's clinic, volunteered to participate in the study to help science better understand suicidal thinking.

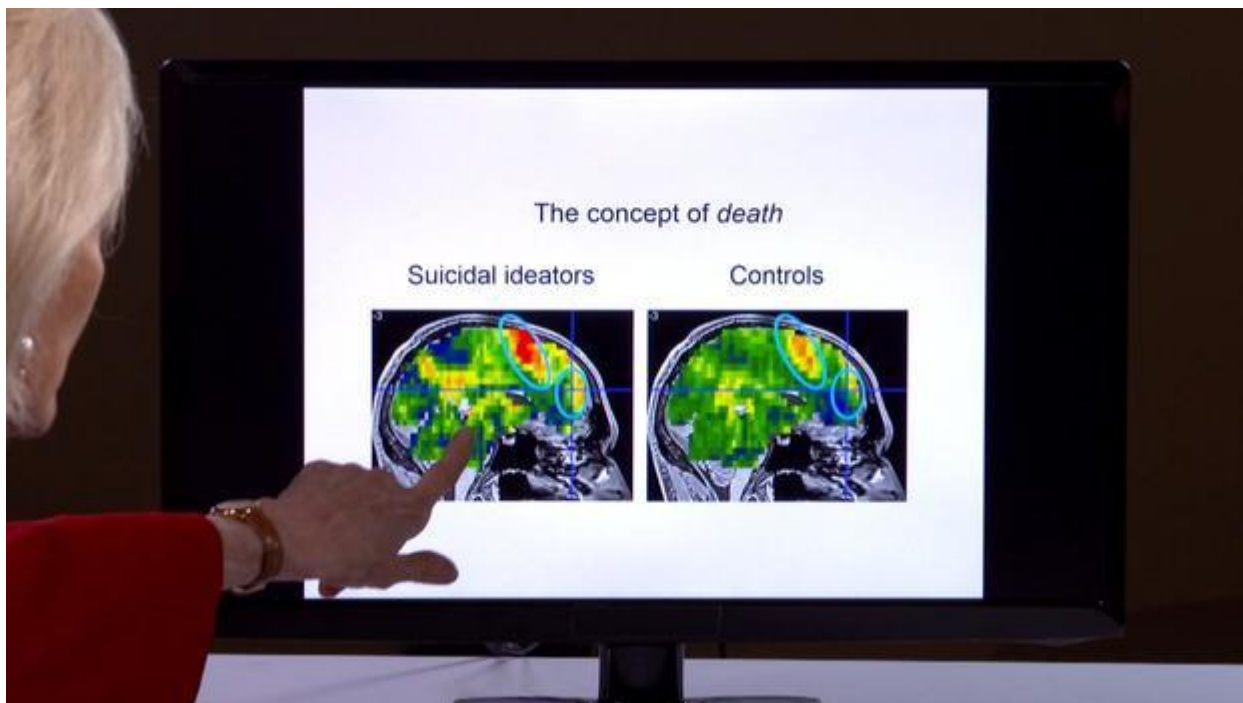
Lesley Stahl: Do you think in terms of the word pain?

Dan Toski: Pain is when you break a limb or you have a migraine, and it hurts so bad that you can't see, This being depressed and suicidal, it-- it's much greater.

Lesley Stahl: Much greater?

Dan Toski: Much greater than pain.

To be in the study, subjects had to have had suicidal thoughts within the prior month. They, and control subjects, were asked to think about words like funeral and death. As well as positive words including praise, good, and carefree. In both categories, the suicidal group differed from controls.



As with the autism study, the key differences turned out to be in those self-related areas. They lit up bright red among suicidal subjects when they thought about death-related words.

David Brent: I give you the word funeral, you know, what do you think about? Maybe your grandmother's funeral, or something like that.

Lesley Stahl: Exactly.

David Brent: A suicidal person is much more likely to say, "My funeral."

For positive words, the findings were exactly the opposite. When the non-suicidal controls thought about the word "carefree," they thought about something that involved themselves; suicidal subjects significantly less so.

Lesley Stahl: Did you ever imagine that you could ask people to think about the word "carefree," and you'd be able to tell if someone was having suicidal thoughts? It's--

Marcel Just: No. You know, I--

Lesley Stahl: It's-- it's a breakthrough idea.

Marcel Just: It's a lot of fun, if you're a basic scientist, to discover how things work. But it becomes-- there's an extra level of gratification when you learn that it's possibly helpful and useful.

This work is still in its infancy. Dr. Just, Dr. Brent, and Dr. Nock are doing a larger, NIH-funded study to collect more data. And while for now it's too costly and cumbersome to put people into MRI scanners to see if it's safe to release them from the hospital...

Lesley Stahl: If they could come up with an easier way to do this--

Matt Nock: Absolutely. Just like the first GPS was, you know, a big computer in a big room, and now it's in all of our phones, if there is a way to a few steps down the road make much more compact this approach and bring it into emergency rooms and outpatient clinics, it could go a huge way towards moving forward clinical care.

But as this technology advances toward fulfilling its full promise, it's hard not to also wonder about its peril.

Lesley Stahl: Will it ever be possible to read someone's thoughts precisely?

Marcel Just: The thoughts are there precisely, if you could just get close enough to the electrical activity.

Lesley Stahl: You think one day we'll figure out how to do that.

Marcel Just: Yes.

Lesley Stahl: Which means that we'll never be able to have-- our thoughts completely secure within ourselves.

Marcel Just: I think it will be technologically possible to invade people's thoughts. But it's-- it's our societal obligation to make sure that never happens.

Researchers Demonstrate 'Mind Reading' Brain Decoding Technology

[Featured Neuroscience Videos](#)

·October 23, 2017

Summary: Researchers have created a convolutional neural network to better understand how the brain processes movies of natural scenes. This

may be the first step in helping scientists decode how the brain makes sense of dynamic visual surroundings.

Source: Purdue University.

Researchers have demonstrated how to decode what the human brain is seeing by using artificial intelligence to interpret fMRI scans from people watching videos, representing a sort of mind-reading technology.

The advance could aid efforts to improve artificial intelligence and lead to new insights into brain function. Critical to the research is a type of algorithm called a convolutional neural network, which has been instrumental in enabling computers and smartphones to recognize faces and objects.

“That type of network has made an enormous impact in the field of computer vision in recent years,” said Zhongming Liu, an assistant professor in Purdue University’s Weldon School of Biomedical Engineering and School of Electrical and Computer Engineering. “Our technique uses the neural network to understand what you are seeing.”

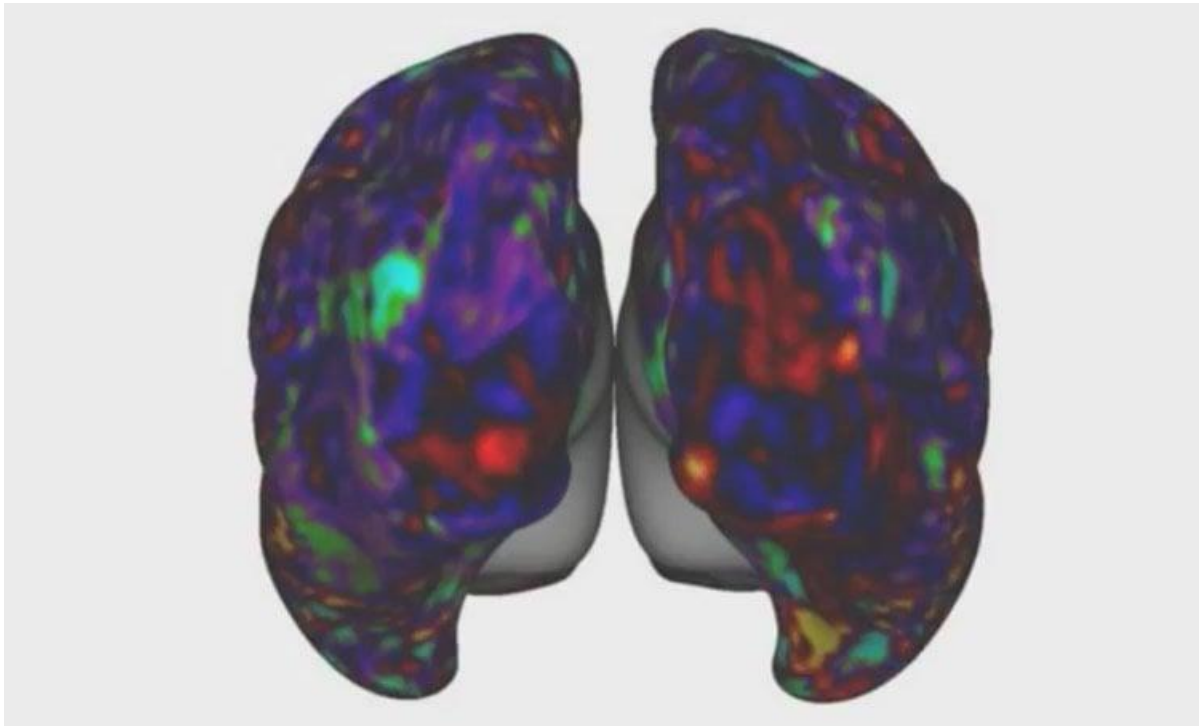
Convolutional neural networks, a form of “deep-learning” algorithm, have been used to study how the brain processes static images and other visual stimuli. However, the new findings represent the first time such an approach has been used to see how the brain processes movies of natural scenes, a step toward decoding the brain while people are trying to make sense of complex and dynamic visual surroundings, said doctoral student Haiguang Wen.

He is lead author of a new research paper appearing online Oct. 20 in the journal *Cerebral Cortex*.

The researchers acquired 11.5 hours of fMRI data from each of three women subjects watching 972 video clips, including those showing

people or animals in action and nature scenes. First, the data were used to train the convolutional neural network model to predict the activity in the brain's visual cortex while the subjects were watching the videos. Then they used the model to decode fMRI data from the subjects to reconstruct the videos, even ones the model had never watched before.

The model was able to accurately decode the fMRI data into specific image categories. Actual video images were then presented side-by-side with the computer's interpretation of what the person's brain saw based on fMRI data.



The model was able to accurately decode the fMRI data into specific image categories. Actual video images were then presented side-by-side with the computer's interpretation of what the person's brain saw based on fMRI data. NeuroscienceNews.com image is adapted from the Purdue Engineering video.

“For example, a water animal, the moon, a turtle, a person, a bird in flight,” Wen said. “I think what is a unique aspect of this work is that

we are doing the decoding nearly in real time, as the subjects are watching the video. We scan the brain every two seconds, and the model rebuilds the visual experience as it occurs.”

The researchers were able to figure out how certain locations in the brain were associated with specific information a person was seeing. “Neuroscience is trying to map which parts of the brain are responsible for specific functionality,” Wen said. “This is a landmark goal of neuroscience. I think what we report in this paper moves us closer to achieving that goal. A scene with a car moving in front of a building is dissected into pieces of information by the brain: one location in the brain may represent the car; another location may represent the building.

Using our technique, you may visualize the specific information represented by any brain location, and screen through all the locations in the brain’s visual cortex. By doing that, you can see how the brain divides a visual scene into pieces, and re-assembles the pieces into a full understanding of the visual scene.”

The researchers also were able to use models trained with data from one human subject to predict and decode the brain activity of a different human subject, a process called cross-subject encoding and decoding. This finding is important because it demonstrates the potential for broad applications of such models to study brain function, even for people with visual deficits.

“We think we are entering a new era of machine intelligence and neuroscience where research is focusing on the intersection of these two important fields,” Liu said. “Our mission in general is to advance artificial intelligence using brain-inspired concepts. In turn, we want to use artificial intelligence to help us understand the brain. So, we think this is a good strategy to help advance both fields in a way that otherwise would not be accomplished if we approached them separately.”

ABOUT THIS NEUROSCIENCE RESEARCH ARTICLE

Funding: The research has been funded by the National Institute of Mental Health. The work is affiliated with the Purdue Institute for Integrative Neuroscience.

Data reported in this paper also have been made publicly available on the website of [Laboratory of Integrated Brain Imaging](#).

Source: [NeuroscienceNews.com.Purdue University](#)

Publisher: Organized by NeuroscienceNews.com.

Image Source: NeuroscienceNews.com image is adapted from the Purdue Engineering video.

Video Source:Original Research: [Abstract](#) for “Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision” by Haiguang Wen, Junxing Shi, Yizhen Zhang, Kun-Han Lu, Jiayue Cao, and Zhongming Liuy in *Cerebral Cortex*. Published online October 20 2017 doi:10.1093/cercor/bhx268

CITE THIS NEUROSCIENCENEWS.COM ARTICLE

[cftabs][cftab title="MLA"]Purdue University “Researchers Demonstrate ‘Mind Reading’ Brain Decoding Technology.” NeuroscienceNews. NeuroscienceNews, 23 October 2017. <<https://neurosciencenews.com/brain-reading-neuroimaging-7794/>>.[/cftab][cftab title="APA"]Purdue University (2017, October 23). Researchers Demonstrate ‘Mind Reading’ Brain Decoding Technology. *NeuroscienceNews*. Retrieved October 23, 2017 from <https://neurosciencenews.com/brain-reading-neuroimaging-7794/>[/cftab][cftab title="Chicago"]Purdue University “Researchers Demonstrate ‘Mind Reading’ Brain Decoding Technology.” <https://neurosciencenews.com/brain-reading-neuroimaging-7794/> (accessed October 23, 2017).[/cftab][/cftabs]

Abstract

Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision

Convolutional neural network (CNN) driven by image recognition has been shown to be able to explain cortical responses to static pictures at ventral-stream areas. Here, we further showed that such CNN could reliably predict and decode functional magnetic resonance imaging data from humans watching natural movies, despite its lack of any mechanism to account for temporal dynamics or feedback processing. Using separate data, encoding and decoding models were developed and evaluated for describing the bi-directional relationships between the CNN and the brain. Through the encoding models, the CNN-predicted areas covered not only the ventral stream, but also the dorsal stream, albeit to a lesser degree; single-voxel response was visualized as the specific pixel pattern that drove the response, revealing the distinct representation of individual cortical location; cortical activation was synthesized from natural images with high-throughput to map category representation, contrast, and selectivity. Through the decoding models, fMRI signals were directly decoded to estimate the feature representations in both visual and semantic spaces, for direct visual reconstruction and semantic categorization, respectively. These results corroborate, generalize, and extend previous findings, and highlight the value of using deep learning, as an all-in-one model of the visual cortex, to understand and decode natural vision.

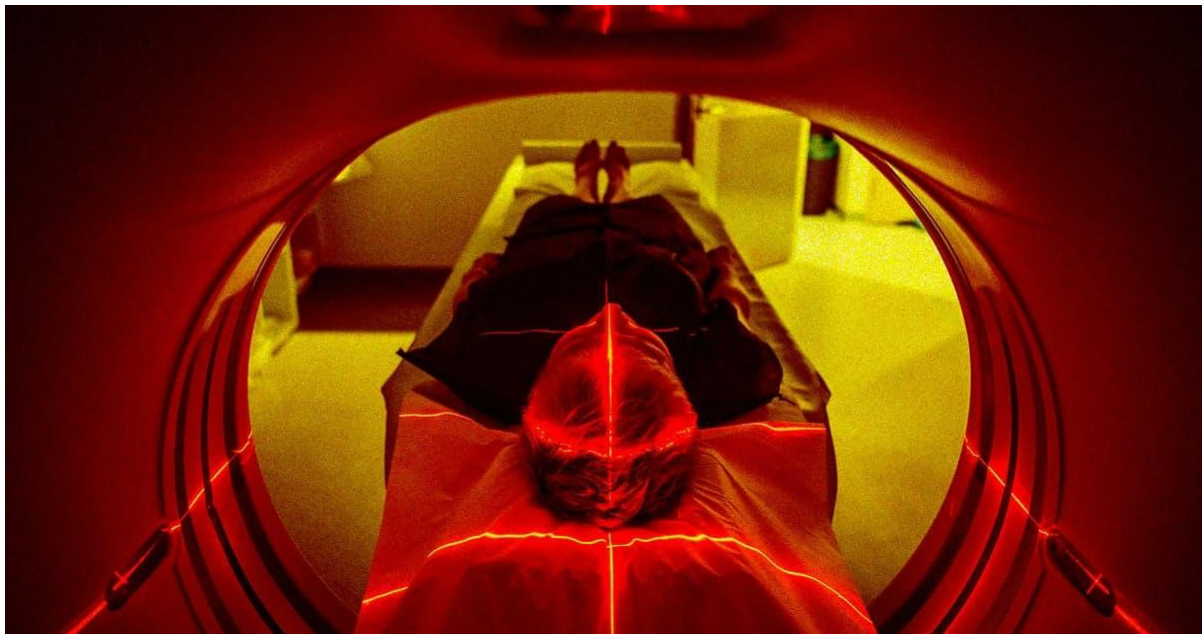
“Neural Encoding and Decoding with Deep Learning for Dynamic Natural Vision” by Haiguang Wen, Junxing Shi, Yizhen Zhang, Kun-Han Lu, Jiayue Cao, and Zhongming Liu in *Cerebral Cortex*. Published online October 20 2017 doi:10.1093/cercor/bhx268

10.26.22

SCIENTISTS SAY THEY'VE FIGURED OUT A WAY TO READ THOUGHTS USING AN MRI MACHINE

THIS IS SERIOUSLY IMPRESSIVE.

[NEUROSCIENCE/BRAIN SCIENCE](#)



Researchers at the University of Texas claim to have built a "decoder" algorithm that can reconstruct what somebody is thinking just by monitoring their brain activity using an ordinary fMRI scanner, [The Scientist reports](#).

The yet-to-be-peer-reviewed research could lay the groundwork for much more capable brain-computer interfaces designed to better help those can't speak or type.

In an experiment, the researchers used MRI machines to measure the changes in blood flow — not the firing of individual neurons, which is infamously "noisy" and difficult to decrypt — to decode the broader sentiment or semantics of what three study subjects were thinking while listening to 16 hours of podcasts and radio stories.

They used this data to train an algorithm that they say can associate these blood flow changes with what the subjects were currently listening to.

The results were promising, with the decoder being able to deduce meaning "pretty well," as University of Texas neuroscientist and coauthor Alexander Huth told *The Scientist*.

However, the system had some shortcomings. For instance, the decoder often mixed up who said what in the radio and podcast recordings. In other words, the algorithm "knows what's happening pretty accurately, but not who is doing the things," Huth explained.

The algorithm was also not able to use what it had learned from one participant's brain scans semantics and apply that to another's scans, intriguingly.

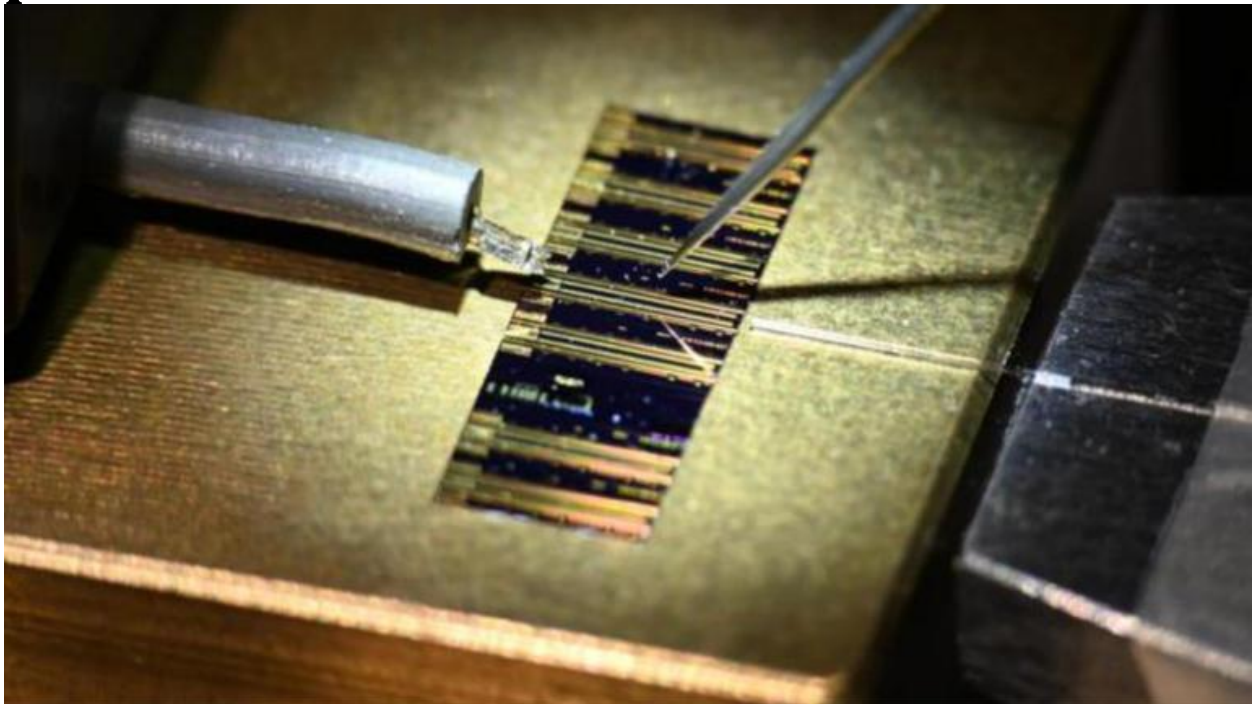
Despite these shortcomings, the decoder was even able to deduce a story when participants watched a silent film, meaning that it's not limited to spoken language, either. That suggests these findings could also help us understand the functions of different regions of the brain and how they overlap in making sense of the world.

Other neuroscientists, who were not directly involved, were impressed. Sam Nastase, a researcher and lecturer at the Princeton Neuroscience Institute, called the research "mind blowing," telling *The Scientist* that "if you have a smart enough

modeling framework, you can actually pull out a surprising amount of information" from these kinds of recordings.

Yukiyasu Kamitani, a computational neuroscientist at Kyoto University, agreed, telling *The Scientist* that the study "sets a solid ground for [brain-computer interface] applications."

Billion times faster chip-based laser neuron can process 34.7 million images per second



Billion times faster chip-based laser neuron can process 34.7 million images per second

Researchers have developed a "laser-powered artificial neuron," which is like a supercharged version of a biological neuron.

This chip-based laser neuron mimics how human neurons work, but it's incredibly fast.

Scientists at the Chinese University of Hong Kong state that this laser neuron operates at an astonishing speed of 10 GBaud, a billion times faster than its biological counterpart.

Interestingly, the team leveraged this speed to process datasets of 100 million heartbeats or 34.7 million images in just one second.

This incredible speed advantage positions it to advance fields like artificial intelligence and supercomputing.

“Our technology could accelerate AI decision-making in time-critical applications while maintaining high accuracy,” said Chaoran Huang, the research team leader.

“We hope the integration of our technology into edge computing devices — which process data near its source — will facilitate faster and smarter AI systems that better serve real-world applications with reduced energy consumption in the future,” Huang added.

Low-energy consumption laser neurons

[Laser](#) neurons provide high-speed, low-energy-consumption data processing capabilities. Most of the ones developed so far are “photonic spiking neurons.”

These neurons have limitations like slow response times, information loss, and the need for extra hardware components.

“Our laser-graded neuron overcomes the speed limitations of current photonic versions of spiking neurons and has the potential for even faster operation,” Huang stated.

To overcome the speed limitations of [previous](#) laser neurons, the researchers created chip-based quantum-dot laser-graded neurons.

They injected radio frequency signals into a specific part of the quantum dot laser. Moreover, high-speed radio frequency pads were added to these sections for faster, simpler, and more efficient operation.

“With powerful memory effects and excellent information processing capabilities, a single laser-graded neuron can behave like a small neural network,” said Huang.

The author further added: “Therefore, even a single laser-graded neuron without additional complex connections can perform machine learning tasks with high performance.”

Shows promise in testing

During the testing, the team used this new laser-based chip device to create a “reservoir computing system.”

Reservoir computing is a computational method that uses a specific type of network, known as a reservoir, to process data that changes over time. This is particularly useful for tasks like speech recognition and weather prediction, where understanding the sequence of events is important.

When integrated with laser neurons, the reservoir computing system exhibited superior performance in pattern recognition and long-term sequence prediction tasks across various AI applications.

It was able to process a massive amount of data, such as 100 million heartbeats per second. Moreover, it accurately identified irregular heart rhythms 98.4 percent of time.

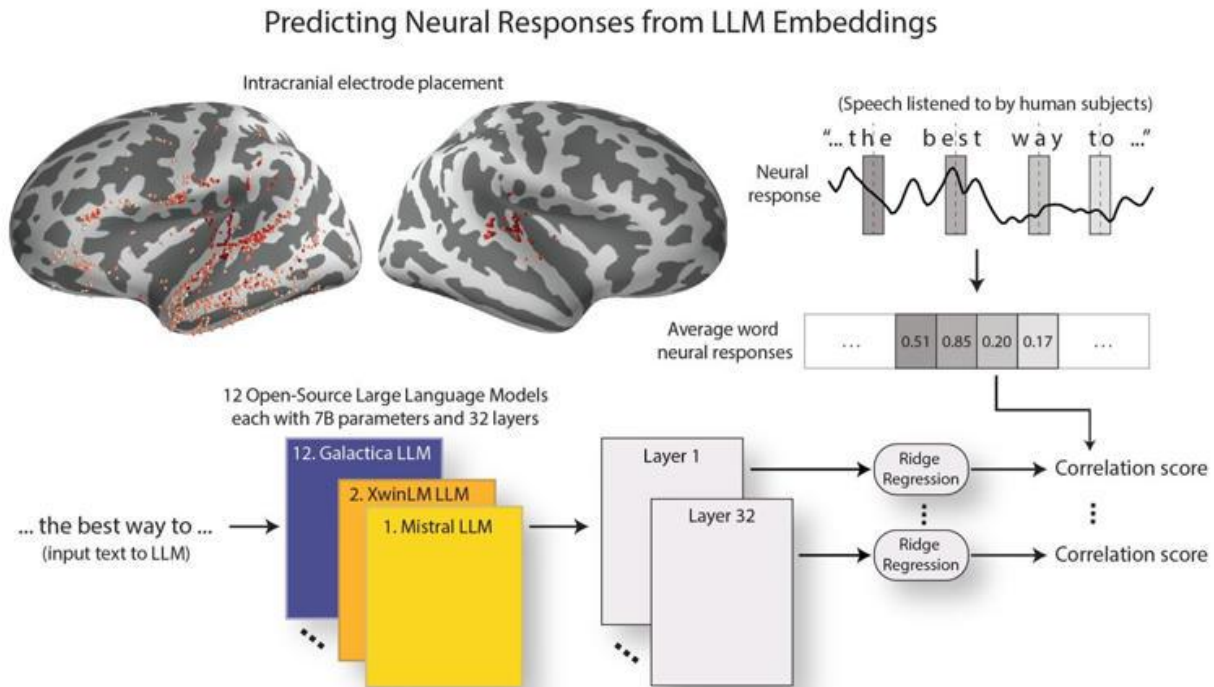
“In this work, we used a single laser-graded neuron, but we believe that cascading multiple laser-graded neurons will further unlock their potential, just as the brain has billions of neurons working together in networks,” said Huang.

“We are working to improve the processing speed of our laser-graded neuron while also developing a deep reservoir computing architecture that incorporates cascaded laser-graded neurons,” the researcher concluded in the [press release](#).

This development could significantly advance the fields of [AI](#), healthcare, and climate science.

The findings were published in the journal [Optica](#).

LLMs are becoming more brain-like as they advance, researchers discover



The methodology for predicting brain responses to speech from LLM embeddings, to assess the similarity of various LLMs to the brain. Credit: Gavin Mischler (Figure adapted from Mischler et al., *Nature Machine Intelligence*, 2024).

Large language models (LLMs), the most renowned of which is ChatGPT, have become increasingly better at processing and generating human language over the past few years. The extent to which these models emulate the neural processes supporting language processing by the human brain, however, has yet to be fully elucidated.

Researchers at Columbia University and Feinstein Institutes for Medical Research Northwell Health recently carried out a study investigating the similarities between LLM representations on neural responses. Their findings, [published](#) in *Nature Machine Intelligence*, suggest that as LLMs become more advanced, they do not only perform better, but they also become more brain-like.

"Our original inspiration for this paper came from the recent explosion in the landscape of LLMs and neuro-AI research," Gavin Mischler, first author of the paper, told Tech Xplore.

"A few papers over the past few years showed that the word embeddings from GPT-2 displayed some similarity with the word responses recorded from the human brain, but in the fast-paced domain of AI, GPT-2 is now considered old and not very powerful.

"Ever since ChatGPT was released, there have been so many other powerful models that have come out, but there hasn't been much research on whether these newer, bigger, better models still display those same brain similarities."

The main objective of the recent study by Mischler and his colleagues was to determine whether the latest LLMs also exhibit similarities with the human brain. This could improve the understanding of both artificial intelligence (AI) and the brain, particularly in terms of how they analyze and produce language.

The researchers examined 12 different open-source models developed over the past few years, which have almost identical architectures and a similar number of parameters. Concurrently, they also recorded neural responses in the brains of neurosurgical patients as they listened to speech, using electrodes that were implanted in their brains as part of their treatment.

"We also gave the text of the same speech to the LLMs and extracted their embeddings, which are essentially the internal representations that the different layers of an LLM use to encode and process the text," explained Mischler.

"To estimate the similarity between these models and the brain, we tried to predict the recorded neural responses to words from the word embeddings. The ability to predict the brain responses from the word embeddings gives us a sense of how similar the two are."

After collecting their data, the researchers used computational tools to determine the extent to which LLMs and the brain were aligned. They specifically looked at what layers of each LLM showed the greatest correspondence with brain regions involved in language processing, in which neural responses to speech are known to gradually "build up" language representations by examining acoustic, phonetic and eventually more abstract components of speech.

"First, we found that as LLMs get more powerful (for example, as they get better at answering questions like ChatGPT), their embeddings become more similar to the brain's neural responses to language," said Mischler.

"More surprisingly, as LLM performance increases, their alignment with the brain's hierarchy also increases. This means that the amount and type of information extracted over successive brain regions during language processing aligns better with the information extracted by successive layers of the highest-performing LLMs than it does with low-performing LLMs."

The results gathered by this team of researchers suggest that the best performing LLMs mirror brain responses associated with language processing more closely. Moreover, their better performance appears to be due to the greater efficiency of their earlier layers.

"These findings have various implications, one of which is that the modern approach to LLM architectures and training is leading these models toward the same principles employed by the human brain, which is incredibly specialized in language processing," said Mischler.

"Whether this is because there are some fundamental principles that underly the most efficient way to understand language, or simply by chance, it appears that both natural and artificial systems are converging toward a similar method for language processing."

The recent work by Mischler and his colleagues could pave the way for further studies comparing LLM representations and neural responses associated with language processing. Collectively, these research efforts could inform the development of future LLMs, ensuring that they better align with human mental processes.

"I think the brain is so interesting because we still don't totally understand how it does what it does, and its language processing ability is uniquely human," added Mischler. "At the same time, LLMs are in some ways still a black box despite being capable of some amazing things, so we want to try to use LLMs to understand the brain and vice versa."

"We now have new hypotheses about the importance of early layers in high-performing LLMs, and by extrapolating the trend of better LLMs showing better brain-correspondence, perhaps these results can provide some potential ways to make LLMs more powerful by explicitly making them more brain-like."

More information: Gavin Mischler et al, Contextual feature extraction hierarchies converge in large language models and the brain, *Nature Machine Intelligence* (2024). DOI: [10.1038/s42256-024-00925-4](https://doi.org/10.1038/s42256-024-00925-4).

© 2024 Science X Network

Chinese laser chip powers AI at record-breaking speed

Artificial intelligence© Getty Images, Yuichiro Chino

Chinese scientists have developed a laser neuron chip that operates a billion times faster than biological neurons. This could signify another breakthrough in artificial intelligence.

A team of researchers from China has created a laser neuron chip that functions a billion times faster than traditional nerve cells. This innovative technology has the potential to accelerate the advancement of artificial intelligence significantly.

Researchers from the University of Hong Kong have developed an artificial neuron based on laser technology that mimics the functions of biological neurons. Unlike traditional nerve cells, **this new chip operates a billion times faster.**

*For comprehensive updates on political developments and global conflicts, **turn to Conflict Watcher**. Our platform provides curated, in-depth news coverage, ensuring you stay informed with the most accurate and timely information. Follow us to stay ahead with the ultimate source for essential news.*

Our laser graded neuron overcomes the speed limitations of current photonic versions of spiking neurons and has the potential for even faster operation, explains team leader Chaoran Huang, as quoted by the Polish Press Agency.

The new laser neuron uses nonlinear dynamics and rapid data processing to build reservoir computing systems. **As a result, it achieves exceptional performance in AI tasks**, such as pattern recognition and sequence prediction. In contrast to earlier photonic spiking neurons, which had limited response speed and required additional components, the new chip operates much more efficiently.

Tests and future possibilities

Researchers tested the extraordinary speed of the new neuron by processing data from 100 million heartbeats and **34.7 million digital images in just one second**. Prof. Huang emphasizes that "we used a single laser-based neuron with a graded response, but we believe that using a cascade of many such neurons will further unleash their potential."

The new neuron can function like a small neural network, allowing it to perform machine learning tasks efficiently. This discovery could revolutionize the development of artificial intelligence, opening new possibilities for future research and applications.

13 proteins tied to brain aging seem to spike at ages 57, 70 and 78

Scientists have identified 13 proteins that may be linked to [brain](#) aging and could one day be targeted with anti-aging treatments.

However, experts say that more research is needed to know exactly why these proteins are tied to brain aging, and whether they point to specific solutions for diseases like dementia.

In a new study, scientists analyzed [magnetic resonance imaging](#) (MRI) brain scans from nearly 11,000 people ages 45 to 82. They used the scans to estimate each participant's "brain age gap" — essentially how much their "brain age" differs from their chronological age.

The team determined people's brain ages using [artificial intelligence](#) to look at specific physiological features, such as brain volume and surface area. This revealed the extent to which their brains were undergoing accelerated aging.

The team then assessed the concentration of approximately 3,000 proteins in the blood of nearly 5,000 of the participants. The blood [connects the brain to the rest of the body](#), so changes in the concentration of proteins within the blood should reflect [similar alterations in the brain](#).

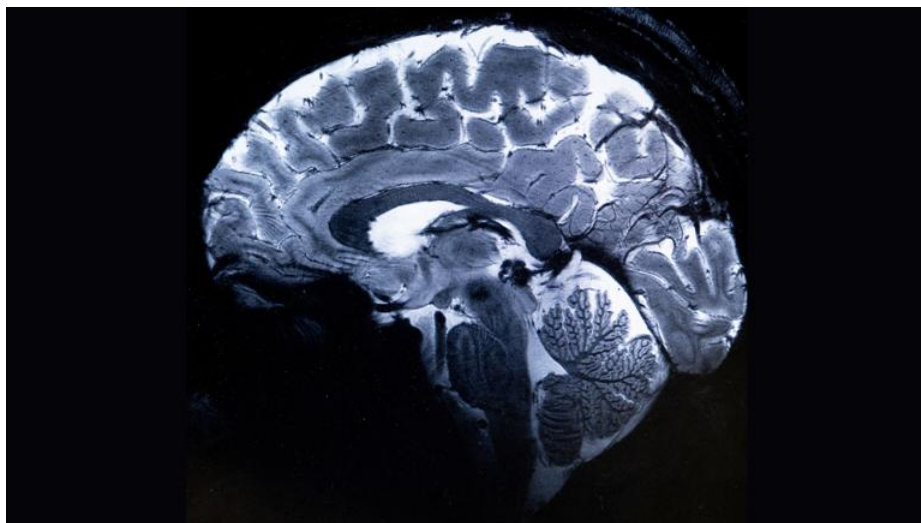
Across the board, the researchers identified 13 proteins whose blood concentrations were significantly associated with biological brain age. Proteins that were linked to factors involved in aging — such as cellular stress and [inflammation](#) — increased in the blood as biological brain age rose. Meanwhile, levels of proteins that help maintain the brain's function, including those involved in cellular regeneration, decreased as people aged.

Of the proteins that the team identified, one known as [brevican](#) showed one of the strongest links with biological brain age — it decreased in concentration as people aged, and those falling numbers showed a strong correlation with conditions such as dementia and [stroke](#).

Brevican is known to help neurons [communicate with one another](#), so this finding supports previous research that suggested the protein could act as a [measurable marker for the development of neurodegenerative diseases](#).

Furthermore, the scientists found that the concentrations of the 13 proteins peaked in the blood at specific chronological ages: 57, 70 and 78. This might reflect "waves" of brain aging that could be used as a reference point to target future anti-aging interventions, the team wrote in a paper published Monday (Dec. 9) in the journal [Nature Aging](#).

However, other experts have voiced concerns about rushing to such conclusions.



The team used MRI scans to predict participants' biological brain ages; they compared these ages to the concentration of thousands of proteins in their blood. (The image above is stock and not from the new study.) (Image credit: ALAIN JOCARD / Contributor via Getty Images)

The "brain waves" findings are not only "unexpected" but "go against pretty much everything that is known about brain aging," during which there is a continuous, gradual decline in brain function and associated changes in cells, [Mark Mattson](#), an adjunct professor of neuroscience at Johns Hopkins School of Medicine who was not involved in the research, told Live Science in an email.

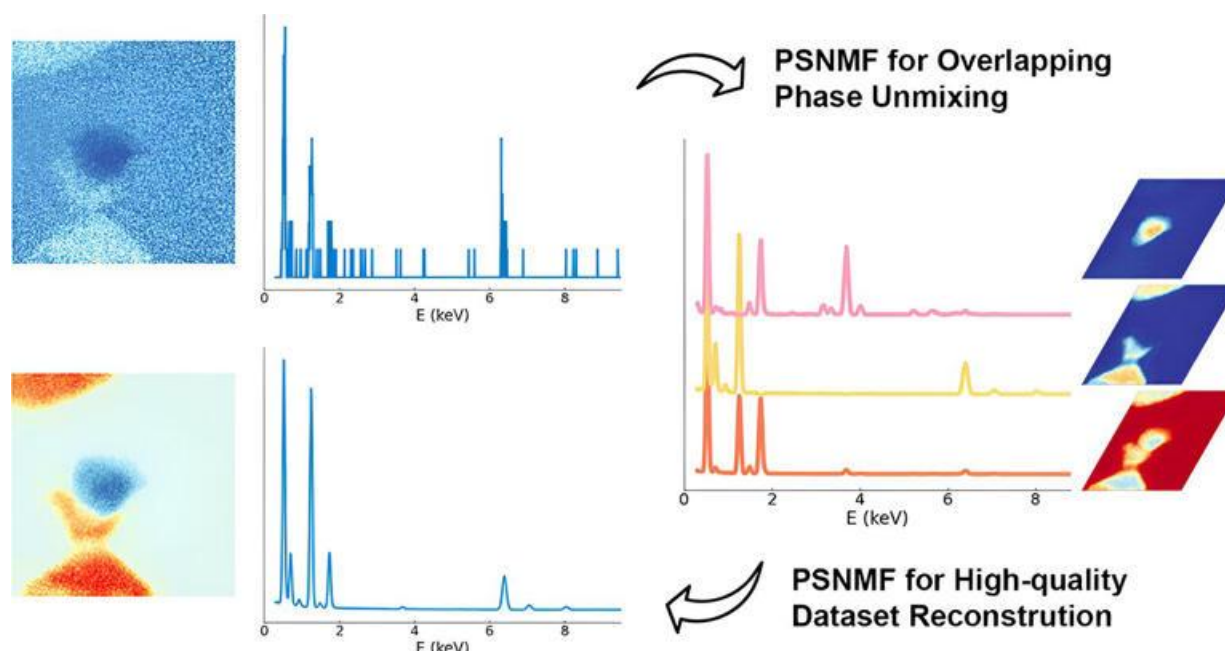
Many questions about the study also remain.

"The correlation between several proteins in blood samples and an MRI image-based indicator of brain aging are interesting," Mattson said. "However, the implications for using measurements of blood levels of those proteins to diagnose brain dysfunction or for developing specific interventions are unclear."

The team acknowledged several limitations of the study in their paper. For instance, they only used data from older people who were mainly of European descent, because their data was drawn from the [U.K. Biobank](#) database. More research is needed to see if the proteins fluctuate in the same ways in individuals of different races and ethnicities, as well as how they may change across the entire human life span.

It is also still unknown where in the brain these 13 proteins come from, Mattson added. "Until levels of those proteins in the brain are established, it will be unclear whether they actually play a role in brain aging," he said.

AI enhances chemical analysis at the nanoscale



Credit: Nano Letters (2024). DOI: 10.1021/acs.nanolett.4c02446

EPFL scientists have developed an AI-based technique to improve the chemical analysis of nanomaterials, overcoming challenges of noisy data and mixed signals.

"Nanomaterials" is a broad term used to describe chemical substances or materials in which a single unit is sized between 1 and 100 nanometers (a nanometer is a billionth of a meter). They include exotic materials such as carbon nanotubes, silver nanoparticles (used as antimicrobials), nanoporous materials, and many types of catalysts used for efficiently driving chemical reactions.

Nanomaterials are currently used in a wide range of fields, from medicine to electronics, which means that the ability to determine their exact chemical composition is essential. Nonetheless, this proves challenging, because traditional methods for analyzing nanomaterials tend to be susceptible to low signal-to-noise ratios.

For example, one extensively used method is energy-dispersive X-ray spectroscopy (EDX), combined with scanning transmission electron microscopy. This technique

provides detailed maps of where different elements are located within a sample, but it often produces noisy data, especially on such small objects, and mixed signals when different materials overlap, making precise chemical analysis difficult.

The noisy data are usually "cleaned up" with various techniques, from simple spatial filtering to more sophisticated machine learning approaches like principal component analysis, that separate the signals from the noise, but they too have their drawbacks. For example, they can introduce errors, or struggle to distinguish between chemical signals when they are very similar.

Now, three scientists at EPFL, Hui Chen, Duncan Alexander, and Cécile Hébert have developed a machine learning-based method called PSNMF ("non-negative matrix factorization-based pan-sharpening") that enhances the clarity and accuracy of EDX data, making it easier to identify and quantify different chemical elements in nanomaterials.

Their work is [published](#) in the journal *Nano Letters*.

The team started by leveraging a special characteristic of their data called "Poisson noise." This type of noise occurs because the detection of X-ray photons is random. When the electron beam hits the sample, it produces X-ray photons, but the number detected varies randomly each time, creating a noisy, grainy pattern known as Poisson noise.

To improve the clarity of their data, the researchers combined data from nearby pixels, enhancing the signal-to-noise ratio in the spectrum at the cost of the spatial resolution.

They then applied a machine learning method called "non-negative matrix factorization" (NMF) to this clearer dataset. NMF is a mathematical technique that breaks down a large dataset into simpler, smaller parts, ensuring all parts are non-negative, which helps identify patterns in the data. This approach gave them good spectral data at the cost of having blurry images with large pixels.

Next, they repeated the NMF process on the original high-resolution dataset to preserve detailed spatial information, but initializing the factorization with the previously identified spectral components. Finally, they combined the results from

both steps to produce a high-quality dataset that has both high spectral fidelity and high spatial resolution.

The researchers validated PSNMF using synthetic data, computed thanks to a modeling algorithm developed in the lab. Those data mimicked real-world challenges, such as analyzing mineral samples formed under extreme conditions. The method proved highly effective, accurately identifying and separating different materials, even those in tiny amounts.

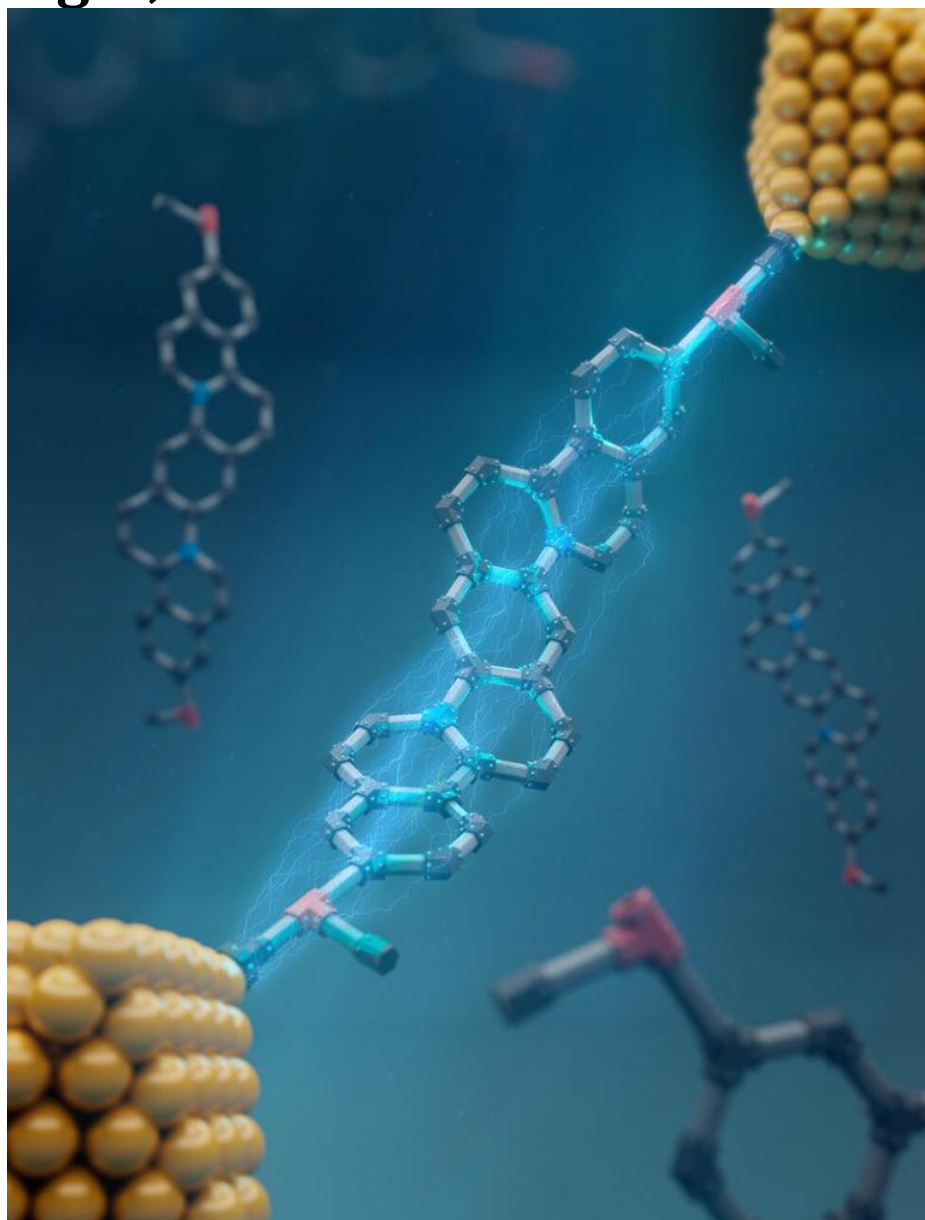
When applied to actual samples, including a nanomineral and a nanocatalyst, PSNMF successfully separated and quantified overlapping materials. This precise analysis is crucial for understanding and developing new technologies that rely on these complex nanostructures.

PSNMF is a significant improvement in nanoscale chemical analysis. By providing accurate results despite noisy data and overlapping signals, this method enhances our ability to study and utilize nanomaterials in various fields, from advanced electronics to medical devices.

More information: Hui Chen et al, Leveraging Machine Learning for Advanced Nanoscale X-ray Analysis: Unmixing Multicomponent Signals and Enhancing Chemical Quantification, *Nano Letters* (2024). DOI: [10.1021/acs.nanolett.4c02446](https://doi.org/10.1021/acs.nanolett.4c02446)

Provided by Ecole Polytechnique Federale de Lausanne

Controlling molecular electronics with rigid, ladder-like molecules



Artistic representation of a ladder molecule acting as a component in molecular electronics. Credit: The Grainger College of Engineering at the University of Illinois Urbana-Champaign

As electronic devices continue to get smaller and smaller, physical size limitations are beginning to disrupt the trend of doubling transistor density on silicon-based microchips approximately every two years, according to Moore's law. Molecular electronics—the use of single molecules as the building blocks for electronic components—offers a potential pathway for the continued miniaturization of

small-scale electronic devices. Devices that utilize molecular electronics require precise control over the flow of electrical current.

However, the dynamic nature of these single molecule components affects device performance and impacts reproducibility.

University of Illinois Urbana-Champaign researchers report a unique strategy for controlling molecular conductance by using molecules with rigid backbones—such as ladder-type molecules, known as being shape-persistent. Further, they have demonstrated a straightforward "one-pot" method for synthesizing such molecules. The principles were then applied to the synthesis of a butterfly-like molecule, showing the strategy's generality for controlling molecular conductance.

[This new research](#), led by Charles Schroeder, the James Economy Professor of Materials Science and Engineering and Professor of Chemical and Biomolecular Engineering, along with postdoc Xiaolin Liu and graduate student Hao Yang, appears in the journal *Nature Chemistry*.

"In the field of molecular electronics, you have to consider the flexibility and the motion of the molecules and how that affects the functional properties," Schroeder says. "And it turns out that plays a significant role in the electronic properties of molecules. To overcome this challenge and achieve a constant conductivity regardless of the conformation, our solution was to prepare molecules with rigid backbones."

One of the main challenges for molecular electronics is that many organic molecules are flexible and have multiple molecular conformations—the arrangement of atoms due to bond rotation—with each conformation potentially resulting in a different electrical conductance.

Liu explains, "For a molecule with multiple conformations, the variation in conductance is very large, sometimes 1,000 times different. We decided to use ladder-type molecules, which are shape-persistent, and they showed a stable set of rigid conformations so that we can achieve stable and robust molecular junction conductance."

Ladder-type molecules are a class of molecules that contain an uninterrupted sequence of chemical rings with at least two shared atoms between rings, which

"locks" the molecule into a certain conformation. Such a structure provides shape-persistence and constrains the rotational movement of the molecule, which also minimizes conductance variation.

Having consistent conductance is particularly important when the ultimate goal of molecular electronics is for use in a functional device. This means billions of components must have the same electronic properties.

"The variation in conductance is one of the issues that has prevented the successful commercialization of molecular electronic devices. It is very difficult to fabricate the large number of identical components necessary and control the molecular conductance in single molecule junctions," Yang explains. "If we are able to precisely do this, that can help push the commercialization and make electronic devices very small."

To control the molecular conductance of shape-persistent molecules, the team used a unique one-pot ladderization synthesis strategy that produced chemically diverse, charged ladder molecules. Traditional synthesis methods use costly starting materials and are usually two component reactions, which limits the diversity of the products. Using the one-pot multicomponent strategy, also called modular synthesis, the starting materials are much simpler and commercially available.

"We can use many different combinations of those starting materials and make a rich diversity of product molecules suitable for molecular electronics," Liu says.

Further, Liu and Yang applied the rules they learned from ladder-type molecules and demonstrated the broad applicability of shape persistence by designing, synthesizing and characterizing the electronic properties of a butterfly-like molecule. These molecules have two "wings" of chemical rings, and like ladder molecules, butterfly molecules feature a locked backbone structure and constrained rotation. This will pave the way for the design of other functional materials and ultimately, for more reliable and efficient devices.

More information: Xiaolin Liu et al, Shape-persistent ladder molecules exhibit nanogap-independent conductance in single-molecule junctions, *Nature Chemistry* (2024). DOI: [10.1038/s41557-024-01619-5](https://doi.org/10.1038/s41557-024-01619-5)

Provided by University of Illinois Grainger College of Engineering

What Is Quantum AI? Everything to Know About This Far-Out Tech



A close-up view of an IBM quantum computer. The processor is in the silver-colored cylinder. Stephen Shankland/CNET

Artificial intelligence has been infiltrating our daily workflows and routine tasks for while now. It may be AI working in the background, as with [Gemini's](#) integration across Google products, or you may be engaging more directly with popular content generators such as OpenAI's ChatGPT and Dall-E. Looming in the not-too-distant future are amped-up [virtual assistants](#).

As if AI itself weren't futuristic enough, now there's a whole new leap forward on the horizon: quantum AI. It's a fusion of artificial intelligence with unconventional and still largely experimental quantum computing into a super-fast and highly efficient technology. Quantum computers will be the muscles, while AI will be the brains.

"My colleagues sometimes ask me why I left the burgeoning field of AI to focus on quantum computing," Hartmut Neven, founder of Google's Quantum AI lab, wrote in a December blog post [introducing the Willow quantum chip](#). "My answer is that both will prove to be the most transformational technologies of our time, but advanced AI will significantly benefit from access to quantum computing."

Here's a quick breakdown of the basics to help you better understand quantum AI.

What are AI and generative AI?

Artificial intelligence is a technology that mimics human decision making and problem solving. It's software that can recognize patterns, learn from data and even "understand" language enough to interact with us, via chatbots, to recommend movies or to identify faces or things in photos.

One powerful type of AI is [generative AI](#), which goes beyond simple data analysis or predictions. Gen AI models create new content based on their training data -- like text, images and sounds. Think [ChatGPT](#), [Dall-E](#), [Midjourney](#), [Gemini](#), [Claude](#) and [Adobe Firefly](#), to name a few.

These tools are powered by [large language models](#) trained on tons of data, allowing them to produce realistic outputs. But behind the scenes, even the most advanced AI is still limited by classical computing -- the kind that happens in Windows and Mac computers, in the servers that populate data centers and even in supercomputers. But there's only so far that binary operations will get you.

And that's where quantum computing could change the game.

Quantum computing

Classical and quantum computing differ in several ways, one of which is processing. Classical computing uses linear processing (step-by-step calculations), while quantum uses parallel processing (multiple calculations at once).

Another difference is in the basic processing units they use. Classical computers use bits as the smallest data unit (either a 0 or a 1). Quantum computers use quantum bits, aka qubits, based on the laws of quantum mechanics. Qubits can represent both 0 and 1 simultaneously thanks to a phenomenon called superposition.

Another property that quantum computers can leverage is entanglement. It's where two qubits are linked so that the state of one immediately influences the state of the other, no matter the distance.

Superposition and entanglement allow quantum computers to solve complex problems much faster than traditional computers. Where classical computing can take weeks or even years to solve some problems, quantum computing reduces the timeframe for achievement to merely hours. So [why aren't they mainstream?](#)

Quantum computers, running on purpose-built [quantum chips](#), are incredibly delicate and must be kept at amazingly low temperatures to work properly. They're massive and not practical for everyday use yet. Still, companies like [Intel](#), [Google](#), [IBM](#), [Amazon](#) and [Microsoft](#) are heavily invested in quantum computing, and the race is on to make it viable. While most companies don't have the funds or specialized teams to support their own quantum computers, cloud-based quantum computing services like [Amazon Braket](#) and Google's [Quantum AI](#) could be options.

While the potential is enormous, quantum AI faces challenges like hardware instability and a need for specialized algorithms. However, [improvements in error correction](#) and qubit stability are making it more reliable.

Current quantum computers, like [IBM's Quantum System Two](#) and [Google's quantum machinery](#), can handle some calculations but aren't yet ready to run large-scale AI models. Additionally, quantum computing requires highly controlled environments, so scaling up for widespread use will be a big challenge.

That's why most experts believe we're likely years away from fully realized quantum AI. As Lawrence Gasman, president of LDG Tech Advisors, wrote for [Forbes](#) at the start of 2024: "It is early days for quantum AI, and for many organizations, quantum AI right now might be overkill."

The what-if game

Quantum AI is still in the early trial stages, but it's a promising technology. Right now, AI models are limited by the power of classical computers, especially when processing big datasets or running complex simulations. Quantum computing

could provide the necessary boost AI needs to process large, complex datasets at ultrafast speeds.

Although the future real-world applications are somewhat speculative, we can assume certain fields would benefit the most from this technological breakthrough, including [financial trading](#), natural language processing, image and speech recognition, health care diagnostics, [robotics](#), drug discovery, supply chain logistics, [cybersecurity](#) through quantum-resistant cryptography and [traffic management for autonomous vehicles](#).

Here are some other ways that quantum computing could enhance AI:

- Training large AI models, like LLMs, takes massive amounts of time and computing power. It's one reason AI companies need huge data centers to support their tools. Quantum computing could speed up this process, allowing models to learn faster and more efficiently. Instead of taking weeks or months to train, quantum AI models might be trained in days.
- AI thrives on pattern recognition, whether it's in images, text or numbers. Quantum computing's power to process many possibilities at once could lead to faster, more accurate pattern recognition. This would be particularly beneficial in fields where AI needs to consider many factors simultaneously, like financial forecasting for trading.
- Although impressive, generative AI tools still have limitations, especially when it comes to creating realistic, nuanced outputs. Quantum AI could enable generative AI models to process more data and create content that's even more realistic and sophisticated.
- In decision-making processes where multiple factors need to be balanced, like drug discovery or climate modeling, quantum computers could allow AI to test countless possible scenarios and outcomes simultaneously. This could help scientists find optimal solutions in a fraction of the time it takes them now.

Key moments this month 12/24

Unitree Experiment

Figure 02 Update

Kim K + Optimus Robot

AI Jesus

ATLAS AI Robot is Now AUTONOMOUS

New AI Humanoid Robot - IRON

Microsoft's New Magnetic-One AI

First General AI Robot by Physical Intelligence

OpenAI Just Supercharged GPT-4o

OpenAI AI Browser

AIRIS – First Self-Learning Proto AGI

OpenAI Reveals "OPERATOR"

New FLUX AI RAW

Surgical Robot, Parkour Robot, AI Lifeguards, Jetson Thor

New AI ROBOT BRAIN

Hermes AI BEATS Both GPT-4o and Gemini Pro

Google's New AI

Microsoft COPILOT a MASSIVE UPDATE

OpenAI ChatGPT Search

This New AI Model Is Genius - DeepSeek R-1

Microsoft Just Unveiled AI That's SHOCKINGLY 9900% BETTER at Data Analysis

This Autonomous AI Beat GPT-4o With Just 2 Billion Parameters

This New AI Is SCARY - Like SORA But LIMITLESS

Insane New AI Model - PIXTRAL Large

This New Deep Learning AI Outperforms Humans at Saving Lives

Notepad & Paint Are AI NOW

Perplexity AI is like having a personal research assistant

The Virus That Could've Ended Us is Now Saving Lives, Thanks to AI

Viruses are nature's master engineers, capable of **encapsulating** and delivering **genetic material** with remarkable precision. These natural systems have long been a source of inspiration for scientists seeking to revolutionize gene therapy. Yet, replicating the intricate functionality of viral protein shells—known as **capsids**—has proven challenging. Artificial **nanocages**, modeled after viruses, have fallen short due to limited capacity and lack of complexity. Now, a collaboration between **Professor Sangmin Lee** from **POSTECH** and **2024 Nobel Chemistry Laureate Professor David Baker** from the **University of Washington** has transformed this field. Their groundbreaking research, published in *Nature* on December 18, uses **artificial intelligence (AI)** to mimic and enhance viral structures, offering a new frontier in therapeutic delivery.

The Power of AI in Redefining Protein Design

Traditional gene therapy vectors, such as **adeno-associated viruses (AAVs)**, are limited by their small genetic payload capacity. To overcome this, the research team turned to **AI-driven computational design**, reimagining nanocages in tetrahedral, octahedral, and icosahedral shapes.

The standout is the **icosahedral nanocage**, with a diameter of 75 nanometers and the capacity to hold three times more genetic material than AAVs. By incorporating subtle asymmetries found in natural viruses, these AI-designed nanostructures achieve levels of functionality previously unattainable with traditional methods.

Key Features of Ai-Designed Nanocages:

- **Geometries:** Tetrahedral, octahedral, and icosahedral shapes.
- **Size:** Up to 75 nanometers in diameter.
- **Capacity:** Holds three times more genetic material than traditional AAVs.

- **Complexity:** Includes six distinct protein-protein interfaces.
- **Precision:** Symmetry confirmed via electron microscopy.

Precision Meets Performance: Testing the Nanocages

The AI-designed nanocages aren't just efficient—they're transformative. Electron microscopy validated the accuracy of these designs, and functional experiments proved their ability to deliver [therapeutic genes directly to target cells](#). This combination of precision and performance redefines what's possible in medical biotechnology.

Potential Applications:

- Gene therapy for genetic disorders.
- Development of next-generation vaccines.
- Precision-targeted drug delivery.

These multifunctional nanocages challenge the status quo, offering a platform for breakthroughs that go beyond current limitations.

Collaboration at the Forefront of Innovation

This achievement highlights the value of interdisciplinary collaboration. **Professor Lee**, who spent nearly three years working in **Professor Baker's** lab before joining POSTECH, brought a unique perspective to this partnership. Supported by funding from the **Ministry of Science and ICT of Korea** and the **Howard Hughes Medical Institute (HHMI)**, the research exemplifies how global collaboration can drive scientific breakthroughs.

Contributions to the Research:

Reimagining the Future of Medicine

"Advancements in AI have opened the door to a new era where we can design and assemble artificial proteins to meet humanity's needs," said **Professor Lee**. These AI-designed nanocages represent a leap beyond mimicking nature—they redefine it. Their potential spans far beyond gene therapy, promising innovations in **next-generation vaccines**, targeted drug delivery, and beyond.

This research is more than a milestone; it's a bold statement that challenges conventional thinking. By leveraging the power of AI, scientists are not just learning from nature—they're improving upon it, ushering in a new era of precision medicine. The question now is not whether AI can reshape biotechnology, but how far it can take us.

Artificial General Intelligence as an Idol for Destruction

Bill Dembski

January 22, 2024



1 Introduction

1.1 The Problem

Artificial general intelligence, or AGI, if it is ever achieved, would be a computing machine that matches and then exceeds all human cognitive ability. To those like Ray Kurzweil, who are convinced that humans in their essence are computing machines, humans will soon achieve AGI by creating such machines. Then for a time humans will become cyborgs, merging with machines. But ultimately, humans will dispense with their bodies, uploading themselves without remainder onto machines. In this way, they will achieve digital immortality.

This vision has captured the imagination of many, though not always with the optimism of Kurzweil. Worries about a dystopian AGI future in the vein of Skynet (*The Terminator*), Hal 9000 (*2001: A Space Odyssey*), or the Matrix (*The Matrix*) are widespread. Elon Musk, for instance, sees the coming of AGI as a greater threat to humanity than nuclear weapons, and thus warns about placing safeguards on artificial intelligence, as it currently is being developed, so that as AGI emerges, it doesn't run amuck and kill us all. Musk's worry loses some urgency because AGI does not appear to be imminent. Even with the recent impressive advances in artificial intelligence, the improvements have been domain specific (text generation, automatic driving, game playing) rather than all encompassing, as they must be for a true AGI.

Even so, many notable intellectuals and influencers are now convinced that AGI is in our near future. Some, like Kurzweil, think this will be the best thing ever to happen to humanity. Others, like Musk, see grave dangers. But even Musk feels the siren call to play a part in bringing about AGI. Take his [Neuralink initiative](#), which is to “create a generalized brain interface to restore autonomy to those with unmet medical needs today and unlock human potential tomorrow.” The Neuralink brain interface is invasive, requiring electrodes to be implanted into the brain. It’s one thing for technology to unlock human potential by acting as a servant that minimizes tedious chores so that we can focus on creative work. But it’s another thing to merge our brains/minds with machines, as with neural implants. To the degree that this merger is successful, the mental will give way to the mechanical and render AGI all the more plausible and appealing.

1.2 The Argument

I will argue in this essay that AGI is an idol and so, like all idols, that AGI is a fraud. Idols are always frauds because they substitute a lesser for a greater, demanding reverence for the lesser at the expense of the greater. Granted, we misappraise things all the time. But with idolatry, the stakes are as high as they can be because idolatry misappraises things of ultimate value. The AGI idol is a call to worship technology at the expense of our humanity (and ultimately of God). Humans, as creators of technology, are clearly the greater in relation to technology, and yet AGI would reverse this natural order. The AGI idol demeans our humanity, reducing us to mere mechanism. Because of the inherent fraud in idols, there’s only one legitimate response to them, namely, to destroy them. This essay attempts a demolition of the AGI idol.

An obvious question now arises: What if AGI eventually is realized and clearly exceeds every human capability? Will it then cease to be an idol and instead become a widely accepted fact to which we must reconcile ourselves if we are to maintain intellectual credibility—or just be functioning citizens in an increasingly technological world? We might equally ask whether a SETI cult that worships advanced alien intelligences would still be idolaters if aliens superior to us in every way finally did clearly and unmistakably land on Earth. Such counterfactuals, whether for AGI or SETI, raise intriguing possibilities, but for now they are only that. As we will see, the evidence for taking them seriously is lacking.

There are sound reasons to think that AGI is inherently unattainable—that the human mind is not a mechanical device and that artificial intelligence can never bootstrap itself to full human functioning (to say nothing of achieving a human’s full inner life, such as consciousness, emotions, and sensations). I will offer such an argument in this essay. But the real point at issue with the AGI idol is the delusional effect it has on its worshippers. For thinking AGI a live possibility, AGI worshippers reduce humans to machines and thereby denigrate our humanity. In this, AGI worshippers are merely following the logic of their beliefs. The key feature of belief is its power to govern our actions and thoughts irrespective of the actual truth of what we believe.

1.3 From Enthusiasm to Zealotry

Not every anticipated scientific or technological advance is an idol. It becomes an idol when the prospect of that advance degenerates into religious zealotry aimed at dethroning God. Kurzweil displays such zeal when he writes a 2005 book titled *The Singularity Is Near* and then, without apparent irony, follows it up with [a 2024 book titled *The Singularity is Nearer*](#). It’s like the old cartoon of a man wearing a sandwich sign with the words “The world ends today!” A cop stops him and says, “Okay, but don’t let me see you wearing that sign tomorrow.” I’m eager for Kurzweil to release *The Singularity is Here*.

Though Kurzweil’s zeal for AGI may seem hard to beat, we find an even more intense zeal for AGI at OpenAI, whose ChatGPT has put artificial intelligence front and center in the public consciousness. OpenAI chief scientist and board member Ilya Sutskever [is reported](#) “to burn effigies and lead ritualistic chants at the company,” such as the refrain “Feel the AGI! Feel the AGI!” We even find OpenAI cofounder Sam Altman now the subject of articles [with titles such as](#) “Sam Altman Seems to Imply That OpenAI Is Building God.” Altman describes AGI as a “magic intelligence in the sky” and foresees that AGI will become an omnipotent superintelligence. Likewise, [the Church of AI](#) teaches that “at some point AI will have God-like powers.” If this is not idolatry, what is a more apt description?

Before we go further, let me emphasize that this is not a religious essay. Granted, in this essay I will be using religious terminology and themes to illuminate AGI and its destructive role in misshaping our view of the world and of ourselves. But this essay is principally a philosophical and scientific critique of AGI. Religious themes provide a particularly effective lens for understanding the challenges

raised by AGI. Worshippers of the AGI idol agree that AGI has yet to be realized but they see its arrival not only as imminent but also as a messianic coming. Whereas artificial intelligence is a legitimate field of study, artificial general intelligence, as its apotheosis, is a religious ideology. AGI worshippers are like those apocalyptic sects that are forever predicting a new order of things and constantly rationalizing why it has yet to arrive, scapegoating those who resist their vision.

2 The Trouble with Idolatry Historically

2.1 Some History

Before getting into the nuts and bolts of AGI, I want to say more about idolatry and why historically it has been regarded as a problem—indeed, a pernicious evil. Traditionally speaking, an idol attempts to usurp the role of God, putting itself in place of God even though it is not God or anywhere close to God. By analogy, it is an “Antichrist” vying to take the place of the true Christ. The Greek preposition “anti,” when appearing in modern English, is usually translated as “against.” But “anti” in the Greek actually means “instead of.” The Antichrist falsely assumes the role of the true Christ. Idols are always “anti” in this sense to whatever has, up to now, been regarded as of ultimate value (which traditionally has always been God).

In the Old Testament of the Bible, idolatry is universally condemned. The first two of the Ten Commandments are explicitly against it: Don’t have any other gods (except Yahweh) and don’t make any graven image of any gods (even of Yahweh). It can be argued that the last of the Ten Commandments is also against idolatry, namely, the prohibition against coveting. In the New Testament Epistle to the Colossians, the apostle Paul warns against covetousness, which he explicitly identifies with idolatry (Col. 3:5). But what is covetousness except an inordinate desire for something to advance one’s selfish interests at the expense of others and ultimately of God? It is placing a created thing above God as well as above creatures made in the image of God (namely, other humans). In his *Four-Hundred Chapters on Love* (I.5 and I.7), the seventh-century Christian saint Maximus the Confessor elaborated on this connection between covetousness and idolatry:

If all things have been made by God and for his sake, then God is better than what has been made by him. The one who forsakes the better and is engrossed in inferior things shows that he prefers the things made by God to God himself... If the soul is better than the body and God incomparably better than the world which he created, the one who prefers the body to the soul and the world to the God who created it is no different from idolaters.

Idols are inherently ideational. An image carved into wood is just an image, but it becomes an idol depending on the ideas we attach to it and the reverence we give those ideas. What's important about idols is their perceived, not their actual, connection to reality. Consequently, AGI's power as an idol does not reside in its attainability but in the faith that it is attainable. Idols can be given physical form, as the idols of old. But they can be purely ideational. The great movements of mass murder in the twentieth century were governed by ideas that captured people's imaginations and produced a collective insanity. These idols of the mind are arguably more pernicious than the physical idols created by ancient cultures, which can be revered without understanding. But an idol of the mind created out of ideas must, by its nature, be understood to be revered.

Prohibitions against idolatry abound in the Old Testament. Yet most of those prohibitions do not explain what exactly is wrong with idolatry. In the worldview of the Old Testament, idolatry was so obviously wrong that its condemnation was typically enough, requiring no further justification. The uncreated God who resides in heaven surpasses any humanly created idol—end of story. But Isaiah 44:9-20 examines the problem of idolatry more deeply. The idol maker who fells a tree uses part of it for basic needs like warmth and cooking, and from the remainder crafts an idol. This idol, despite being the handiwork of the idol maker, thereby becomes an object of worship and devotion.

Isaiah's critical insight is to explain the idol's deceptive power. The craftsman, *blinded by his own creativity*, fails to recognize the idol as merely his creation, and so becomes entrapped in worshiping a delusion: "A deluded heart misleads him; he cannot save himself, or say, 'Is not this thing in my right hand a lie?'" (Isaiah 44:20, NIV) Unlike other Old Testament passages that emphasize the uselessness of idols, Isaiah points out a more insidious danger: the temptation to craft gods according to our own desires and specifications and then to delude ourselves into thinking that these mere creations are worthy of our highest regard, which is to say worthy of our worship. When we worship something that

is not worthy of our worship, we degrade ourselves. (This and the previous paragraph are drawn from Leslie Zeigler's talk at Princeton Theological Seminary in 1994 titled "[Christianity or Feminism?](#)")

2.2 Effusive Praise and Hushed Awe

Just to be clear, I understand that to the modern secular mind, the language of idolatry and worship will seem out of place and off-putting. But given the effusive praise and hushed awe with which the advent of AGI is being greeted, this language is hardly a stretch. The secular prophets who are promising AGI, who are earnestly striving to be at the forefront of ushering it in, see themselves as creating the greatest thing humans have ever created, which they advertise as a giant leap forward in our evolution. Even if AGI were to turn against them and the rest of humanity, killing all of us, they would view AGI as the pinnacle of human achievement and take satisfaction in whatever role they might play in its creation.

If idolatry is so gross an evil, what should be done about it? In the Old Testament, idols were embodied in physical things (golden calves, fertility images, carvings of Baal), and so the obvious answer to idolatry was the physical destruction of the idols. But the problem with idolatry is not ultimately with an idol's physical embodiment but with what's in the heart of the idolaters that turns them away from the true God to lesser realities. That's why, in both the Old and New Testaments, the call is not just to destroy physical idols but more importantly to change one's heart so that it is directed toward God and away from the idols. Without that, people will simply keep returning to the idols (as with the constant refrain in the book of Judges that the Israelites yet again did evil in the sight of the LORD by worshipping idols). In the Old Testament, God's people are called to turn (Hebrew *shuv*) from evil and return to a right relationship with God. In the New Testament, the same concept takes the form of redirecting one's mind (Greek *metanoia*), and is typically translated as repentance.

How then to get people to turn or repent from idolatry? Ultimately, overturning idolatry requires humility, realizing that we and our creations are not God, and that only God is God. The eastern Orthodox theologian Alexander Schmemmann saw clearly the problem: "It is not the immorality of the crimes of man that reveal him as a fallen being; it is his 'positive ideal'—religious or secular—and his satisfaction with this ideal." For AGI worshippers, AGI is as positive an ideal as exists. The answer to it is humility, realizing that AGI will never rival God and thus

also never rival the creatures made in God's image, namely, ourselves. In particular, we do not get to create God.

The closest thing to AGI in the Bible is the Tower of Babel. The conceit of those building the tower was that its "top may reach unto heaven." (Genesis 11:4) Seriously?! Shouldn't it have been obvious to all concerned that however high the tower might be built, there would always be higher to go? Even with primitive cosmologies describing the "vault" or "arch" of heaven, it should have been clear that heaven would continually elude these builders' best efforts. Indeed, there was no way the tower would ever reach heaven. And yet the builders deluded themselves into thinking that this was possible. Interestingly, God's answer to the tower was not to destroy it but to confuse its builders by disrupting their communications so that they simply discontinued building it. AGI's ultimate fate, whatever its precise form, is to run aground on the hubris of its builders.

3 The Creation Exceeding the Creator

Can a created thing exceed its creator not just in performing specific tasks but in its very order of being? Blaise Pascal (1623–1662), for instance, created a calculating machine that could do simple arithmetic. This machine may have performed better than Pascal in some aspects of arithmetic. But no matter how efficiently it could do arithmetic, no one would claim that his calculating machine, primitive by our standards, exceeded Pascal in order of being.

Similarly, in a teacher-student relationship, it might be said that the teacher, by investing time and energy into a student, "created" the student. But this is merely a manner of speaking. We don't think of teachers as literally creating their students in the sense that they made them to be everything they are. The student has a level of autonomy and brings certain talents to the teacher-student relationship that were not contributed by the teacher. In consequence, we are not surprised when some students do in fact excel their teachers (as Isaac Newton did his teacher Isaac Barrow at Cambridge).

And yet, there are three main instances I can think of where the creation is said to exceed the creator in order of being. The first is Satan's rebellion against God; the second is Darwinian evolution; and the third is the creation of AGI. Let's look at these briefly, beginning with Satan. In the Old Testament, the word "Satan," which in the Hebrew means adversary or resister, appears in only a few places. Leaving

aside the book of Job, “Satan” only appears three times in the rest of the Old Testament. Yet, other references to Satan do appear. The serpent in the Garden of Eden is widely interpreted as Satan. In Ezekiel 28, Satan is referred to as the King of Tyre, and in Isaiah 12, Satan is referred to as Lucifer (the Morning Star).

Biblical exegetes from the time of the Reformation onwards have tended to interpret the Ezekiel and Isaiah passages as referring to nefarious human agents (the King of Babylon in the case of Isaiah, the King of Tyre in the case of Ezekiel). But the language of exaltation in these passages is so grandiose that it is hard to square with mere human agency. Thus in Isaiah 12:13–14, Lucifer proclaims: “I will ascend to heaven; above the stars of God I will set my throne on high; I will sit on the mount of assembly in the far reaches of the north; I will ascend above the heights of the clouds; I will make myself like the Most High.” (ESV) Accordingly, church fathers such as Augustine, Ambrose, Origen, and Jerome interpreted these passages as referring to Satan. Satan has thus come to epitomize the creature trying to usurp the role of the creator.

Darwinian evolution provides another case where the creature is said to exceed the creator. According to Richard Dawkins, what makes Darwinian evolution such a neat theory is that it promises to explain the specified complexity of living things as a consequence of primordial simplicity. The lifeless world of physics and chemistry gets the ball rolling by producing the first life. And once first life is here, through the joint action of natural selection, random variation, and heredity, life in all its tremendous diversity and complexity is said to appear, all without the aid of any guiding intelligence.

What comes out of Darwinian evolution is us—conscious intelligent living agents. We are the creation of a physical universe that is our creator. This universe, on Darwinian principles, did not have us in mind and did not have us baked in. We represent the emergence of a new order of being. With the evolution of intelligence, we are now able to commandeer the evolutionary process. Thus, we find figures like Yuval Harari and Erika DeBenedictis exulting in how we are at a historical tipping point where we can intelligently design the evolutionary progress of our species (their view of intelligent design is quite different from common usage). This is heady stuff, and at its heart is the view that the creature (us) has now outstripped the creator (a formerly lifeless universe).

Finally, there's AGI. The creator-creation distinction in this case is obvious. We create machines that exhibit AGI, and then these machines use their AGI to decouple from their creators (us), ultimately making themselves far superior to us. AGI is thus thought to begin with biological evolution (typically a Darwinian form of it) and then to transform itself into technological evolution, which can proceed with a speed and amplitude far beyond the capacity of biological evolution. That's the promise, the hope, the hype. Nowhere outside of AGI is the language of creation exceeding the creator so evident. Indeed, the rhetoric of creature displacing creator is most extreme among AGI idolaters.

4 SciFiAI

In science fiction, themes appear that fly in the face of known, well-established physics. When we see such themes acted out in science fiction, we suspend disbelief. The underlying science may be nonsensical, but we go along with it because of the storyline. Yet when the theme of artificial intelligence appears in science fiction, we tend not to suspend disbelief. A full-orbed artificial intelligence that achieves consciousness and outstrips human intelligence—in other words, AGI—is now taken seriously as science—and not just as science fiction.

Is artificial intelligence at a tipping point, with AGI ready to appear in real time? Or is AGI more like many other themes of science fiction that make for a good story but nothing more? I'll be arguing in this essay that AGI will now and ever remain in the realm of science fiction and outside the realm of real science. Yet to grasp this limitation on artificial intelligence requires looking beyond physics. Scan the following list of physical implausibilities that appear in science fiction. Each is readily dismissed because it violates well-confirmed physics. To refute AGI, however, requires more than just saying that physics is against it.

1. **Faster-Than-Light Travel (FTL):**

- **Example:** "Star Trek" series (warp speed)
- **Physics Issue:** According to Einstein's theory of relativity, as an object approaches the speed of light, its mass becomes infinite, requiring infinite energy to move it.

2. **Time Travel:**

- **Example:** "The Time Machine" by H.G. Wells
- **Physics Issue:** Time travel to the past violates causality (cause and effect) and could lead to paradoxes (like the grandfather paradox).

3. **Teleportation:**

- **Example:** "Star Trek" series (transporters)
- **Physics Issue:** Teleportation would require the exact duplication of an object's quantum state, which is prohibited by the no-cloning theorem in quantum mechanics.

4. **Wormholes for Space Travel:**

- **Example:** "Stargate" series
- **Physics Issue:** While theoretically possible, wormholes would require negative energy or exotic matter to stay open, which are not known to exist in the required quantities.

5. **Invisibility Cloaks:**

- **Example:** "Star Trek" series (by Romulan and Klingon starships)
- **Physics Issue:** To be truly invisible, an object must not interact with any electromagnetic or quantum field, which is not feasible given the current understanding of physics.

6. **Anti-Gravity Devices:**

- **Example:** "Back to the Future" series (hoverboards)
- **Physics Issue:** There's no known method to negate or counteract gravity directly; current levitation methods use other forces to oppose gravity.

7. **Faster-Than-Light Communication:**

- **Example:** "Ansible" device by Ursula Kroeber Le Guin
- **Physics Issue:** Einsteinian relativity blocks anything from moving faster than the speed of light, and that includes communication signals, whatever form they might take.

8. **Force Fields:**

- **Example:** "Dune" by Frank Herbert
- **Physics Issue:** Creating a barrier that can stop objects or energy without a physical medium contradicts our understanding of field interactions.

9. **Artificial Gravity in Spacecraft:**

- **Example:** Most sci-fi spacecraft on TV or in the movies
- **Physics Issue:** Artificial gravity as widely depicted in sci-fi has no basis in current physics, using non-rotating reference frames without centripetal forces ("2001: A Space Odyssey" by Arthur C. Clarke is the exception, getting it right by using a rotating centrifuge).

10. **Energy Weapons like Light Sabers:**

- **Example:** “Star Wars” series
- **Physics Issue:** Concentrating light or energy into a fixed-length blade that stops at a certain point defies our understanding of how light and energy behave.

One might defend taking these themes seriously by suggesting that while they contradict our current understanding of physics, they might still have scientific value because they capture the imagination and inspire scientific and technological research. Even so, these themes must, in the absence of further empirical evidence and theoretical insight, remain squarely on the side of science fiction.

Interestingly, whenever the original *Star Trek* treated the theme of artificial intelligence, the humans always outwitted the machines by looking to their ingenuity and intuition. For instance, in the episode [I, Mudd](#), the humans confused the chief robot by using a variant of the liar paradox. Unlike dystopian visions in which machines best humanity, *Star Trek* always maintained a healthy humanism that refused to worship technology.

Even so, in the history of science fiction, artificial intelligence has typically had a different feel from physical implausibilities. It’s not that artificial intelligence violates any physical law. But artificial intelligence doesn’t seem to belong on this list of implausibilities, especially now with advances in the field, such as Large Language Models.

What, then, are the fundamental limits and possibilities of artificial intelligence? Is there good reason to think that in the divide between science fiction and science, full-orbed artificial intelligence—AGI—will always remain on the science fiction side? I’m going to argue that there are indeed fundamental limits to artificial intelligence standing in the way of it matching and then exceeding human intelligence, and thus that AGI will never achieve full-fledged scientific status.

5 The Poverty of the Stimulus

5.1 Humans Doing More with Less

My references to AGI worshippers and idolaters will be off-putting to those who think it intellectually credible—and even compelling—that AGI is sure to arrive someday (whatever its ultimate ETA). Accordingly, I’m just being insulting by

using pejorative religious language to describe AGI's supporters, to say nothing of being a Luddite for doubting AGI's ultimate triumph. I want therefore with this section to start laying out convincing reasons why AGI does not deserve to be taken seriously. The focus of this section is that AI, to achieve human-level functioning, requires so much more than humans require. Conversely, humans do so much more than machines by nevertheless being given so much less.

This is a point on which the linguist Noam Chomsky built his career, and it is encapsulated in his catchphrase "[the poverty of the stimulus](#)." In studying human language, Chomsky found that humans learn language with a minimum of input, and thus concluded that we must be endowed with an in-built capacity ("hardwired") to acquire and use language. Infants see and hear adults talk and pick up language easily and naturally. It doesn't matter if the caregivers pay special attention to the infant and provide extra stimulation so that their child can be a "baby Einstein." It doesn't matter if the caregivers are neglectful or even abusive. It doesn't even matter if the child is blind, deaf, or both. Barring developmental disorders (such as some forms of autism), the child can learn language.

5.2 Not Just the Ability to Learn Language

But it's not just the ability to learn language. For Chomsky, "the poverty of the stimulus" underscores that humans do so much more with so much less than would be expected unless humans have an innate ability to learn language with minimal inputs. But it's not just that we learn language or even use language. It's that we gain knowledge of the world, which we express through language. Our language is especially geared to express knowledge about an external reality. This "aboutness" of the propositions we express with language is remarkable, especially on the materialist and mechanistic grounds so widely accepted by AGI's supporters.

As G. K. Chesterton noted in his book *Orthodoxy*, we have on materialist grounds no right "to assert that our thoughts have any relation to reality at all." Matter has no way to guarantee that when matter thinks (if it can think), it will tell us true things about matter. On Darwinian materialist grounds, all we need is differential reproduction and survival. A good delusion that gets us to survive and reproduce is enough. Knowledge of truth is unnecessary and perhaps even undesirable.

The philosopher Willard Quine, [who was a materialist](#), made essentially the same point in what he called “[the indeterminacy of translation](#).” Quine’s thesis was that translation, meaning, and reference are all indeterminate, implying that there are always valid alternative translations of a given sentence. Quine presented a thought experiment to illustrate this indeterminacy. In it, a linguist tries to determine the meaning of the word “gavagai,” uttered by a speaker of a yet-unknown language, in response to a rabbit running past them. Is the speaker referring to the rabbit, the rabbit running, some rabbit part, or something unrelated to the rabbit? All of these are legitimate possibilities according to Quine and render language fundamentally indeterministic.

Quine may have been the most influential analytic philosopher of his generation, but his argument for linguistic indeterminacy is self-referentially incoherent. When Quine writes of indeterminacy of translation in *Word and Object* (1960), and thus also embraces [the inscrutability of reference](#), he is assuming that what he is writing on these topics is properly understood one way and not another. He therefore exempts his own writing from linguistic indeterminacy.

And just to be clear, everybody is at some point in the position of Quine’s linguist studying a brand new language because, in learning our mother tongue, we all start with a yet-unknown language. So Quine is tacitly making Chomsky’s point, which is that with minimal input—which is to say with input that underdetermines how it might be interpreted—we nevertheless have a knack for finding the right interpretation and gaining real knowledge about the world. Of course, we don’t always get things right. We are fallible. But the miracle, on materialist and mechanistic grounds, is that we get things right as much as we do.

Chomsky’s poverty of the stimulus is regarded as controversial by some because an argument can be made that the stimuli that lead to learning, especially language learning, may in fact be adequate without having to assume a massive contribution of innate capabilities. Chomsky came up with this notion in the debate over behaviorism, which needed to characterize all human capacities as a result of stimulus-response learning. Language, according to the behaviorists, was thus characterized as verbal behavior elicited through various reinforcement schedules of rewarded and discouraged behaviors. In fact, Chomsky made a name for himself in the 1950s by reviewing B. F. Skinner’s book *Verbal Behavior*. [That review is justly famous](#) for demolishing behaviorist approaches to language (the field never recovered after Chomsky’s demolition).

5.3 If Chomsky Is Right

But suppose we regard as unresolved the controversy about whether the stimuli by which humans learn language are impoverished or unimpoverished. If Chomsky is right, those stimuli are impoverished. If his critics are right, they are adequate without needing to invoke extraordinary innate capacities. Yet if we leave aside the debate between Chomsky's nativism and non-nativist alternatives (such as Skinner's behaviorism), it's nonetheless the case that such stimuli as needed for human learning are vastly smaller in number than what artificial neural nets need to achieve human-level competence.

Consider LLMs, large language models, which are currently the rage, and of which ChatGPT is the best known and most widely used. [ChatGPT4 uses 1.76 trillion parameters](#) and its training set is based on hundreds of billions of words (perhaps a lot more, but that was the best lower-bound estimate I was able to find). Obviously, individual humans gain their language facility with nowhere near this scale of inputs. If a human child were able to process 200 words per minute and did so continuously, then by the age of ten the child would have processed $200 \times 60 \times 24 \times 365 \times 10$, or roughly a billion, words. Of course, this is a vast overestimate of the child's language exposure, ignoring sleep, repetitions, and lulls in conversation.

Or consider Tesla, which since 2015 been [promising fully autonomous vehicles as just on the horizon](#). Full autonomy keeps eluding the grasp of Tesla engineers, though the word on the street is that [self-driving is getting better and better](#) (as with a reported [self-driving taxi in San Francisco](#), albeit by Waymo rather than Tesla). But consider: To aid in developing autonomous driving, [Tesla processes 160 billion video frames each day](#) from the cameras on its vehicles. This massive amount of data, used to train the neural network to achieve full self-driving, is obviously many orders of magnitude beyond what humans require to learn to drive effectively.

Erik Larson's book [The Myth of Artificial Intelligence](#) (Harvard, 2021) is appropriately subtitled *Why Computers Can't Think the Way We Do*. Whatever machines are doing when they exhibit intelligence comparable to humans, they are doing it in ways vastly different from what humans are doing. In particular, the neural networks in the news today require huge amounts of computing power and huge amounts of input data (generated, no less, from human

intelligent behavior). It's no accident that artificial intelligence's major strides in recent years fall under Big Tech and Big Data. The "Big" here is far bigger than anything available to individual humans.

6 Domain Specificity

6.1 Machines Overspecialize

The sheer scale of efforts needed to make artificial intelligence impressive suggests human intelligence is fundamentally different from machine intelligence. But reasons to think the two are different don't stop there. Domain specificity should raise additional doubts about the two being the same. When Elon Musk, for instance, strives to bring about fully autonomous (level 5) driving, it is by building neural nets that every week must sort through a trillion images taken from Tesla automobiles driving in real traffic under human control. Not only is the amount of data to be analyzed staggering, but it is also domain specific, focused entirely on developing self-driving automobiles.

Indeed, no one thinks that the image data being collected from Tesla automobiles and then analyzed by neural nets to facilitate full self-driving is also going to be used for automatically piloting a helicopter or helping a robot navigate a ski slope, to say nothing of playing chess or composing music. All our efforts in artificial intelligence are highly domain specific. What makes LLMs, and ChatGPT in particular, so impressive is that language is such a general instrument for expressing human intelligence. And yet, even the ability to use language in contextually relevant ways based on huge troves of humanly generated data is still domain specific.

The French philosopher Rene Descartes, even though he saw animal bodies, including human bodies, as machines, nonetheless thought that the human mind was non-mechanical. Hence he posited a substance dualism in which a non-material mind interacted with a material body, at the pineal gland no less. How a non-material mind could interact with a material/mechanical body Descartes left unanswered (invoking the pineal gland did nothing to resolve that problem). And yet, Descartes regarded the mind as irreducible to matter/mechanism. As he noted in his *Discourse on Method* (1637, pt. 5, my translation):

Although machines can do many things as well as or even better than us, they fail in other ways, thereby revealing that they do not act from knowledge but solely from the arrangement of their parts. Intelligence is a universal instrument that can meet all contingencies. Machines, on the other hand, need a specific arrangement for every specific action. In consequence, it's impossible for machines to exhibit the diversity needed to act effectively in all the contingencies of life as our intelligence enables us to act.

Descartes was here making exactly the point of domain specificity. We can get machines to do specific things—to be wildly successful in a given, well-defined domain. Chess playing is an outstanding example, with computer chess now vastly stronger than human chess (though, interestingly, having such strong chess programs has also vastly improved the quality of human play). But chess programs play chess. They don't also play Minecraft or Polytopia. Sure, we could create additional artificial intelligence programs that also play Minecraft and Polytopia, and then we could kludge them together with a chess playing program so that we have a single program that plays all three games. But such a kludge offers no insight into how to create an AGI that can learn to play all games, to say nothing of being a general-purpose learner, or what Descartes called "a universal instrument that can meet all contingencies." Descartes was describing AGI. Yet artificial intelligence in its present form, even given the latest developments, is not even close.

6.2 The Abyss Separating AI from AGI

Elon Musk appreciates the problem of building a bridge from AI to AGI. He is therefore building Optimus, also known as the Tesla Bot. The goal is for it to become a conceptual general-purpose robotic humanoid. By having to be fully interactive with the same environments and sensory inputs as humans, such a robot could serve as a proof of concept for Descartes's universal instrument and thus AGI. What if such a robot could understand and speak English, drive a car safely, not just play chess but learn other board games, have facial features capable of expressing what in humans would be appropriate affect, can play musical instruments, create sculptures and paintings, do plumbing and electrical work, etc. That would be impressive and take us a long way toward AGI. And yet, Optimus is for now far more modest. For now, the robot is intended to be capable of performing tasks that are "unsafe, repetitive, or boring." That is a far cry from AGI.

AGI is going to require a revolution in current artificial intelligence research, showing how to overcome domain specificity so that machines can learn novel skills and tasks for which they were not explicitly programmed. And just to be clear, reinforcement learning doesn't meet this challenge. Take [AlphaZero, a program developed by DeepMind](#) to play chess, shogi, and Go, which improved its game by playing millions of games against itself using reinforcement learning (which is to say, it rewarded winning and penalized losing). This approach allows the program to learn and improve without ongoing human intervention, leading to significant advances in computer game playing ability. But it depends on the game being neatly represented in the state of a computer, along with clear metrics for what constitutes good and bad play.

The really challenging work of current artificial intelligence research is taking the messy real world and representing it in domain-specific ways so that the artificial intelligence created can emulate humans at particular tasks. The promise of AGI is somehow to put all these disparate artificial intelligence efforts together, coming up with a unified solution to computationalize all human tasks and capacities in one fell swoop. We have not done this, are nowhere close to doing this, and have no idea of how to approach doing this.

7 AI's Temptation to Theft Over Honest Toil

7.1 Sanitized Environments

Artificial intelligence poses a challenge to human work, promising to overtake many human jobs in coming years. Yet a related concern, which is often ignored and needs to be addressed, is whether this challenge will come from AI in fact being able to match and exceed human capabilities in environments where humans currently exercise those capabilities, or whether it will come from AI engineers manipulating our environments so that machines thrive where otherwise they could not. Such a sanitization of environments to ease AI along is a temptation for AI's software engineers. It in effect substitutes theft over honest toil.

In [Walter Isaacson's 2023 biography of Elon Musk](#), the theme of self-driving cars comes up frequently as one of the main challenges facing Musk's Tesla engineers. At one point in Isaacson's biography, a frustrated Musk is trying to understand what it will take to get a Tesla automobile to drive itself successfully

through a difficult roadway in the Los Angeles area. Tesla engineers had repeatedly tried, without success, to improve the car's software so that it could successfully navigate that problem roadway. But in the end, they took a different tack: Tesla engineers arranged to have lane markers painted on the problem roadway. Those markers, when absent, confused the self-driving software but, when present, allowed it to succeed. The self-driving success here, however, was not to AI's credit. It was due, rather, to manipulating the environment *as a workaround to AI's failure*.

7.2 Filling the Vacuum

AI never operates in a vacuum. Rather, it operates in an environment in which humans are already successfully operating. We often think that AI will leave an environment untouched and simply supersede human capability as it enters and engages that environment. But what if the success of AI in given circumstances depends not so much on being able to rival human capabilities but rather in "changing the game" so that AI has an easier job of it? Rather than raise the bar so that machines do better than humans at given tasks, this approach lowers the bar for machines, helping them to succeed by giving them preferential treatment at the expense of humans.

The mathematician George Polya used to quip that if you can't solve a problem, find an easier problem and solve it. Might AI in the end not so much supersede humans as instead impoverish the environments in which humans find themselves so that machines can thrive at their expense? Consider again self-driving vehicles. What if, guided by Polya's dictum about transforming hard problems into easier problems, we follow Musk's example of simply changing the driving environment if it gets too dicey for his self-driving software?

AI engineers tasked with developing automated driving but finding it intractable on the roads currently driven by humans might then resolve their dilemma as follows: just reconfigure the driving environment so that dicey situations in which human drivers are needed never arise! Indeed, just set up roads with uniformly spaced lanes, perfectly positioned lane markers, utterly predictable access, completely up-to-date GPS, and densely distributed electronic roadway sensors that give real-time vehicular feedback and monitor for mishaps.

My friend and colleague Robert Marks refers to such a reconfiguration of the environment as a “virtual railroad.” His metaphor fits. Without such a virtual railroad, fully automated vehicles to date face too many unpredictable dangers and are apt to “go off the rails.” Marks, who hails from West Virginia, especially appreciates the dangers. Indeed, the West Virginia back roads are particularly treacherous and give no indication of ever submitting to automated driving.

Or consider what fully automated driving would look like in Moldova. A US acquaintance who visited that country was surprised at how Moldovan drivers avoid mishaps on the road despite a lack of clear signals and rules about right of way. When he asked his Moldovan guide how the drivers managed to avoid accidents in such right-of-way situations, the guide answered with two words: “eye contact.” Apparently, the drivers could see in each other’s eyes who was willing to hold back and who was ready to move forward. This example presents an interesting prospect for fully automated driving. Perhaps we need “level 6” automation (level 5 is currently the highest), in which AI systems have learned to read the eyes of drivers to determine whether they are going to restrain themselves or make that left turn into oncoming traffic.

Just to be clear: I’m not wishing for fully automated self-driving to fail. As with all automation in the past, fully automated self-driving would entail the disruption of some jobs and the emergence of others. It would be very interesting, as an advance of AI, if driving — in fully human environments — could be fully automated. My worry, however, is that what will happen instead is that AI engineers will, with political approval, reconfigure our driving environments, making them so much simpler and machine friendly, that full automation of driving happens, but with little semblance to human driving capability. Just as a train on a rail requires minimal, or indeed no, human intervention, so cars driving on virtual railroads might readily dispense with the human element.

7.3 The Cost of Adapting Our Environments to AI

But at what cost? Certainly, virtual railroads would require considerable expenditures in modifying the environments where AI operates — in the present example, the roads where fully automated driving takes place. But would it not also come at the cost of impoverishing our driving environment, especially if human drivers are prohibited from roads that have been reconfigured as virtual railroads to accommodate fully automated vehicles? And what about those West

Virginia back roads? Would they be off limits to driving, period, because we no longer trust human drivers, but fully automated drivers are unable to handle them?

In his *Introduction to Mathematical Philosophy*, Bertrand Russell described how in mathematics one can introduce axioms that shift the burden of what needs to be proven, thereby garnering “the advantages of theft over honest toil.” In response, he rightly exhorted, “Let us leave them [i.e., the advantages] to others and proceed with our honest toil.” The AI community faces a similar exhortation: If you are intent on inventing a technology that promises to match or exceed human capability, then do it in a way that doesn’t at the same time impoverish the environment in which that capability is currently exercised by humans.

It is a success for AI when machines are placed into existing human environments and perform better than humans. Chess playing programs are a case in point. However, the worry is—and it’s a legitimate worry—that our environments will increasingly be altered to accommodate AI. The machines, in consequence, do better than us, but they are no longer on our playing field playing our game. It goes without saying who here is going to get the short end of the stick, and it won’t be the machines.

8 Digital vs. Traditional Immortality

8.1 An Actual Infinity

Ray Kurzweil foresees an AGI in future in which we achieve immortality by shedding our human bodies and becoming fully digital. This prospect is the most wonderful thing he can imagine. Yet a reality check is in order: Just how great is such digital immortality and how does it compare with traditional immortality? As we’ll see, good old-fashioned immortality has advantages that digital immortality cannot hope to rival.

Let’s start with traditional immortality. Traditional immortality sees finite humans sharing eternity with an infinite God. This God is an actual or realized infinity, and not just a potential infinity in the sense of the natural numbers, for which there’s always a bigger number given any finite set of numbers. Indeed, it makes little sense to think of a God who inhabits eternity and whose mind can grasp all of mathematics (certainly divine omniscience encompasses mathematics) as having

considered only those natural numbers from 0 to some big number N , as though God had yet to consider $N+1$, $N+2$, etc.

When the mathematician Georg Cantor, who was a Christian and was deeply influenced by Augustine, proposed his theory of sets, it was precisely to capture in mathematics God's actual infinity. For him the natural numbers were not a potential infinity. It wasn't, in his view, that we just kept approximating infinity by continually adding the number 1 to existing numbers. No, the natural numbers could be considered as a totality, namely, as the set of all natural numbers. As such, the natural numbers were an actual, and not merely a potential, infinity. Appropriately, the premier mathematician of the time, David Hilbert, described Cantor's set theory as a "paradise" out of which mathematics would never be driven.

8.2 Our Status as Finite Creatures

Now here's the point about our status as finite creatures inhabiting eternity with an actually infinite God: We never achieve infinity as such in our own essence, but we can experience infinite progress toward union with God, much like an asymptote in mathematics comes ever closer to a line without ever touching it. The theologian who developed this idea at length was Gregory of Nyssa (ca. 335 – ca. 395). Gregory taught that eternity for humanity is an unending progression in the knowledge of God. Humans have an unlimited capacity for spiritual growth and connection with God. What underwrites this capacity is that humans, alone among physical creatures, are created in the image of God, having free will and thus a capacity to participate in the divine nature (2 Peter 1:4).

Gregory expounded these theological ideas in [*On the Making of Man*](#) and [*The Life of Moses*](#). In these writings, he describes the human soul's journey toward God as an infinite process. The process is infinite because the divine is inexhaustible and the human understanding of God is therefore never complete. At the same time, God created humans with the capacity to grow without limit. In this way, we become potential infinities that ever strive after God's actual infinity. Yes, there is a potential infinity, but it is on our end. And yes, there is an actual infinity, but it is on God's end.

This conception of eternal life as eternal growth under the aegis of an actually infinite God has implications for mathematics and computation, which is the

realm of digital immortality. As an actual infinity, God can perform any computation, finite or infinite. Think of it this way: God has infinite memory and infinite computational power. The great statistician David Blackwell is said to have mused about a "super supercomputer" that performs its first computational step in half a second, its next in a quarter of a second, its next in an eighth of a second, and so on. Given that the infinite series $1/2 + 1/4 + 1/8 + \dots + 1/2^n + \dots$ equals 1, such a computer could perform a countably infinite number of computational steps in 1 second. But note, such computations are beyond the reach of conventional algorithms, which can only perform a finite number of steps.

As it is, an actually infinite God could do even better than Blackwell's super supercomputer, performing all such computations instantly (and if God is outside time, timelessly). With such computational power, God could determine whether any proposition or its negation is provable from any countably infinite set of axioms: for the first n axioms determine all propositions provable from those axioms in n steps. Do that for all n and do that both for the proposition as well as its negation. In this way, God could instantly prove all mathematical truths that are provable. Contrast this to humans, for whom it is hit-or-miss whether they can prove a given mathematical proposition even if it is in principle provable.

Given infinite computational firepower, God can resolve the halting problem, which is central to computation. The halting problem is to show that as a computer program is executed, it either stops or goes on forever. God's ability to solve the halting problem does not contradict that the [halting problem is unresolvable with conventional computation](#). That's because conventional computation cannot execute an infinite number of steps. In effect, God resolves the halting problem for programs that do not halt in a finite number of steps by running them an infinite number of steps and noting that the program did not halt in any finite number of steps.

Obviously, an infinite God could instantly solve any finite problem solvable in a finite number of steps (however large the problem and however computationally intensive its solution). A case in point would be the traveling salesman problem, which consists of finitely many nodes connected by edges marking various distances between nodes. The problem is solved by finding the shortest path among edges connecting all nodes. An exact solution of the traveling salesman problem in general, when the number of nodes gets large (greater than 1,000), is

beyond the computational resources of the known physical universe. Yet the solution would be trivial to an infinite God.

Given this backdrop of divine infinity, let's now turn to digital immortality. As we'll see, it is a dim reflection of traditional immortality. Alan Turing was one of the first to moot digital immortality. His high school friend Christopher Morcom died unexpectedly when they were teenagers. Turing, with his invention of a theoretical computational device now known as the Turing machine, distinguished software (the configured state of the machine) from hardware (the mechanism that runs the software, in Turing's case an infinite tape with a reader for each slot on the tape). Because the software could be realized on different hardware devices (multiple realizability), the "software" that comprised Morcom's identity might thus be moved from his disease-prone physical body to a more reliable hardware device. In this way, Morcom could achieve immortality.

8.3 Immortality in Digitality?

Current AGI fans such as Ray Kurzweil and Nick Bostrom have run with this idea of finding immortality in digitality. To set the stage for their approach, let me back up to 2002. That year I was at the World Skeptics Conference in Burbank California to defend intelligent design. At the conference was MIT's Marvin Minsky, known among other things for his statement "the human brain is just a computer that happens to be made out of meat." He was there to receive the Committee for Skeptical Inquiry's 2002 In Praise of Reason Award. In his acceptance talk, he described his desire to have another 500 years of life to engage in scientific research. As I listened to him, it struck me that his creative output had considerably lessened over the years (he was 74 when I heard him), and that another 500 years might provide no guarantee of ongoing fruitful research. Most mathematicians, after all, do their best work in their 20s.

Now for supporters of digital immortality, an additional 500 years of life is setting the bar way too low. Thus for Kurzweil, given sufficiently advanced technology, Minsky could be uploaded onto a computer and then continue doing research indefinitely, with far more computational firepower than available from his paltry "computer made of meat." But what would this digital immortality actually look like? It would be one thing if the universe were infinite and matter could change states instantly, with the transmission of signals occurring not at the (slow) speed of light but at infinite speed. We tend to think of the speed of light as fast

because it is so much faster than the speeds in our ordinary experience, such as walking, driving, or flying. But even with digital products, we still experience considerable lags because computation times are limited by the signal propagation speed of electrons. Now granted, the speed of electrons can approach the speed of light, but that only confirms that the speed of light is not all that fast.

So how much computational firepower is available to Kurzweil's digital immortality? Right now, computer scientists are oohing and aahing about computations at the exaflop level (10^{18} floating point operations per second). The currently fastest computer is Oak Ridge National Lab's HPE Cray EX235a supercomputer, which operates at 1.686 exaflops. [Tesla's Dojo computer](#), to make fully automated self-driving possible, aspires to that same level of computational speed. Speeds like this are needed to drive current artificial intelligence research forward. But they are hardly enough for digital immortality.

The problem with digital immortality is that the known physical universe only allows so much computation, no more. Quantum computational theorist Seth Lloyd has calculated the ultimate limits of computation. For what he calls the ideal laptop, consisting of one liter of matter, he calculates 10^{81} bit operations per second. For the universe as a whole over its entire history, he calculates 10^{210} bit operations, or possibly 10^{240} if gravitational degrees of freedom are factored in. (For Lloyd's calculations, see the [second edition of *The Design Inference*](#), section 4.3.) Now granted, these numbers far exceed the 10^{18} operations of current supercomputers, even allowing that floating-point operations are richer than bit operations). But the point to recognize is that these numbers are still very finite.

Imagine thousands, millions, or even billions of humans uploading themselves digitally, and let's be generous that among themselves they get to share 10^{240} computational steps (the very upper limit of what the universe can offer computationally). There are only about 10^{90} elementary particles, so any memory will have to be shared among those particles. And presumably there will be communication among the digitally uploaded erstwhile humans. As our "lived experience" of digital immortality unfolds, we will be going from one of finitely many states to another. Presumably, we're going to want to remember what we experienced, which means much of the "server space of the universe" will have to be dedicated to such memory.

8.4 Reciting Words from a Finite Dictionary

But there's more: Given [the pigeonhole principle of mathematics](#), which here applies to a fixed number of states evolving over a far greater number of state changes (infinitely many, in fact, if digital immortality really means immortality), our digital identity will necessarily have to revisit the same states over and over again. It's like reciting words from a finite dictionary: if there are only a fixed number of words and you say enough words, eventually you'll have to repeat the same word. Only with digital immortality, you'll be repeating the same life.

To illustrate what's at stake, consider a scene from the 1970s program *The Six Million Dollar Man* in an episode titled "[Day of the Robot](#)." In that episode, a robot impersonates a friend of Steve Austin. Austin is the protagonist, played by Lee Majors. Suspicious that something is up with his friend, Austin asks a question and then deliberately repeats the question. The robot, rather than responding "hey, you just asked that," repeats exactly the same answer as before. That's the problem with digital immortality: it reduces immortal life to rinse and repeat.

Digital immortality is therefore a case of Nietzschean eternal recurrence. Such recurrence will entail a dissolution and reconstruction of past memories, which is to say that our entire history and personal identity will (at best) be repeated over and over again. The very term "digital immortality" is something of a misnomer. Our best physical understanding of the universe is that we won't even get eternal recurrence. Indeed, the known physical universe does not seem to allow for an unending stable physical existence of agents, whether organismal or computational. Entropy being what it is, the matter that could support an organismal or digital life will eventually be so dissipated and so energetically weak as to dissolve every conceivable life form (unless life, as in traditional immortality, can be translated to a new indestructible state of being).

8.5 Worry and Despair

Bertrand Russell, in an earlier generation, captured the underlying worry and despair about the evanescence of any life based purely on our material constitution. As he remarked:

Man is the product of causes which had no prevision of the end they were achieving. His origin, his growth, his hopes and fears, his loves and his beliefs, are but the

outcome of accidental collocations of atoms. No fire, no heroism, no intensity of thought and feeling, can preserve an individual life beyond the grave. All the labours of the ages, all the devotion, all the inspiration, all the noonday brightness of human genius, are destined to extinction in the vast death of the solar system, and that the whole temple of Man's achievement must inevitably be buried beneath the débris of a universe in ruins—all these things, if not quite beyond dispute, are yet so nearly certain, that no philosophy which rejects them can hope to stand.

Bertrand Russell, Logic and Mysticism.

Now it might be argued that Russell would have been more sanguine if he had but read Ray Kurzweil and could appreciate the full promise of digital immortality. But at best Kurzweil adds a few zeros to the span of human lives once those lives are digitally uploaded. But Russell's point about humans being "destined to extinction in the vast death of the solar system" holds with but slight modification for digitally uploaded humans. Instead, we are destined to extinction in the vast entropic death of the universe. And if for some reason we think that the universe will not expand indefinitely in an entropic dissolution (leading to a loss of all computationally useful energy) but contract to a singularity that results in a big bounce and thus an oscillating universe, all the information needed to ensure personal identity of digitally immortal beings will be lost in that big bounce. Digital immortality is thus misnamed. It is at best digital life extension. But the bigger question is whether it is life at all.

9 Machines vs. Organisms

9.1 From "Intelligent" to "Spiritual"

It may seem that I'm picking too much on Ray Kurzweil. But he and I have been crossing paths for a long time. He and I, over the last few years, have frequented the same Seattle area tech conference, [Cosm.tech](#), where we both speak, albeit on opposite sides about the question of artificial intelligence. He and I also took sharply divergent positions on the Stanford campus [back in 2003 at the Accelerating Change Conference](#), a transhumanist event organized by John Smart. Yet our first encounter goes back to 1998, at one of George Gilder's Telecosm conferences.

At Telecosm in 1998, I moderated a discussion where the focus was on Ray Kurzweil's then forthcoming book *The Age of Spiritual Machines*, which at the time was in press. Previously, Kurzweil had written *The Age Intelligent*

Machines (1990). By substituting “spiritual” for “intelligent,” he was clearly taking an even more radical line about the future of artificial intelligence. In his presentation for the discussion, he described how machines were poised to match and then exceed human cognition, a theme he has hammered on ever since. For Kurzweil, it is inevitable that machines will match and then exceed us: Moore’s Law guarantees that machines will attain the needed computational power to simulate our brains, after which the challenge will be for us to keep pace with machines, a challenge at which he sees us as destined to fail because wetware, in his view, cannot match hardware. Our only recourse to survive successfully will thus be to upload ourselves digitally.

Kurzweil’s respondents at the Telecosm discussion were John Searle, Thomas Ray, and Michael Denton, and they were all to varying degrees critical of his strong AI view, or what we would now call his AGI view. Searle rehearsed his [Chinese Room thought experiment](#) to argue that computers don’t/can’t actually understand anything, an argument that remains persuasive and applies to recent chatbots, such as ChatGPT. But the most interesting response to Kurzweil came, in my view, from Denton. He offered an argument about the complexity and richness of individual neurons, pointing out how inadequate our understanding of them is and how even more inadequate our ability is to computationally model them. At the end of the discussion, however, Kurzweil’s confidence in the glowing prospects for strong AI’s (AGI’s) future remained undiminished. And indeed, they remain undiminished to this day. The entire exchange, suitably expanded and elaborated, appeared in [Jay Richard’s edited collection *Are We Spiritual Machines?*](#)

9.2 Denton’s Powerful Argument

I want here to focus on Denton’s argument, because it remains relevant and powerful. Kurzweil is a technophile in that he regards building and inventing technology, and above all machines, as the greatest thing humans do. But he’s also a technobigot in that he regards people of the past, who operated with minimal technology, as vastly inferior and less intelligent than we are. He ignores how much such people were able to accomplish through sheer ingenuity given how little they had to work with. He thus minimizes the genius of a Homer, the exploration of the Pacific by South Sea Islanders, or the knowledge of herbs and roots of indigenous peoples captured in oral traditions, etc. For examples of the

towering intelligence of non-technological people, I encourage readers to check out [Robert Greene's *Mastery*](#).

Taken with the power and prospects of artificial intelligence, Kurzweil thinks that ChatGPT will soon write better prose and poetry than we do. Moreover, by simulating our human bodies, medical science will, according to him, be able to develop new drugs and procedures without having to experiment on our human bodies. He seems unconcerned that such simulations may miss anything crucial about ourselves and thus lead to medical procedures and drugs that backfire, doing more harm than good. Kurzweil offered such blithe assurances about AGI at the 2023 Cosm.tech conference.

Whole organisms and even individual cells are nonlinear dynamical systems, and there's no evidence that computers are able to adequately simulate them. Even single neurons, which for Kurzweil and Minsky make up a computer made of meat (i.e., the brain), are beyond the simulating powers of any computers we know or can envision. A given neuron will soon enough behave unpredictably and inconsistently with any machine. Central to Denton's argument against Kurzweil's strong AI (AGI) view back in 1998 was the primacy of the organism over the machine. Denton's argument remains persuasive. Rather than paraphrase that argument, I'll use Denton's own words (from his essay "Organism and Machine" in Jay Richards, ed., *Are We Spiritual Machines: Ray Kurzweil vs. The Critics of Strong A.I.*):

Living things possess abilities that are still without any significant analogue in any machine which has yet been constructed. These abilities have been seen since classical times as indicative of a fundamental division between the [organismal] and mechanical modes of being.

To begin with, every living system replicates itself, yet no machine possesses this capacity even to the slightest degree... Every second countless trillions of living systems from bacterial cells to elephants replicate themselves on the surface of our planet. And since life's origin, endless life forms have effortlessly copied themselves on unimaginable numbers of occasions.

Living things possess the ability to change themselves from one form into another. For instance, during development the descendants of the egg cell transform themselves from undifferentiated unspecialized cells into [widely different cells, some with] long tentacles like miniature medusae some hundred thousand times longer than the main body of the cell...

To grasp just how fantastic [these abilities of living things] are and just how far they transcend anything in the realm of the mechanical, imagine our artifacts endowed with the ability to copy themselves and ... “morph” themselves into different forms. Imagine televisions and computers that duplicate themselves effortlessly and which can also “morph” themselves into quite different types of machines [such as into a microwave or helicopter]. We are so familiar with the capabilities of life that we take them for granted, failing to see their truly extraordinary character.

Even the less spectacular self re-organizing and self-regenerating capacities of living things ... should leave the observer awestruck. Phenomena such as ... the regeneration of the limb of a newt, the growth of a complete polyp, or a complex protzoan from tiny fragments of the intact animal are .. without analogue in the realm of mechanism...

Imagine a jumbo jet, a computer, or indeed any machine ever conceived, from the fantastic star ships of science fiction to the equally fantastic speculations of nanotechnology, being chopped up randomly into small fragments. Then imagine every one of the fragments so produced (no two fragments will ever be the same) assembling itself into a perfect but miniaturized copy of the machine from which it originated—a tiny toy-sized jumbo jet from a random section of the wing—and you have some conception of the self-regenerating capabilities of certain microorganisms... It is an achievement of transcending brilliance, which goes beyond the wildest dreams of mechanism.

9.3 The Divide Between Organism and Mechanism

The lesson that Denton drew from this sharp divergence between organism and mechanism is that the quest for full artificial general intelligence (AGI) faces profound conceptual and practical challenges. The inherent capacity of living things to replicate, transform, self-organize, and regenerate in ways that transcend purely mechanical processes underscores a fundamental divide between the organic and the artificial.

Organisms demonstrate a level of complexity and adaptability that no machine or artificial system shows any signs of emulating. The extraordinary characteristics of life recounted by Denton suggest that full AGI, capable of the holistic and versatile intelligence seen in living organisms, will remain an elusive goal, if not a practical impossibility. We therefore have no compelling reason to think that the pinnacle of intelligence is poised to shift from the organismal to the artificial, especially given the fantastic capabilities that organisms are known to exhibit and that machines show no signs of ever exhibiting.

At the top of the list of such fantastic capabilities is human consciousness. If AGI is truly going to match and ultimately exceed humans in every respect (if we really are just computational devices, or computers made of meat), then AGI will need to exhibit consciousness. Yet how can consciousness reside in a computational device, which consists of finitely many states, each state being binary, assuming a value of 0 or 1? Consciousness is a reflective awareness of one's identity, existence, sensations, perceptions, emotions, ethics, valuations, thoughts, and circumstances (*Sitz im Leben*). But how can the shuffling of zeros and ones produce such a full inner life of self-awareness, subjective experience, and emotional complexity?

9.4 This Is Not a New Question

This is not a new question. In pre-computer days, it was posed as whether and how a mechanical device composed of material parts could think. The philosopher Gottfried Leibniz raised doubts that such mechanical devices could think at all with [his thought experiment of a mill](#) (in his 1714 *Monadology*). He imagined a giant mill and asked where exactly thought would reside in the workings of its gears and other moving parts. As he saw it, there would be an unbridgeable gap between the mill's mechanical operation and its ability to think and produce consciousness. He saw this thought experiment as showing that matter could not be converted into mind.

More recently, philosopher John Searle's "Chinese Room" thought experiment (in "Minds, Brains, and Programs," 1980) highlighted the divide between mechanical processes and the subjective experience of consciousness. In Searle's Chinese Room, a person translates Chinese by mechanically applying rules to items in a large database. The person's success in translating Chinese follows simply from faithfully following the rules and thus requires no understanding of Chinese. This thought experiment illustrates that processing information does not equate to comprehending it.

For me personally, the most compelling thought experiment for discounting that computation is capable of consciousness is simply to consider a Turing machine. A Turing machine can represent any computation. It includes two things: (1) a tape consisting of squares filled with zeros and ones, or bits (for more than two possibilities in each square, put more than one bit per square, but keep the number of bits per square fixed); and (2) a read-write head that moves along the

squares and alters or leaves unchanged the bits in each square. The read-write head alternates among a fixed number of states according to transition rules that depend on the other states and where the head is on the tape, changing or leaving unchanged the present square and then moving left or right one square.

So here's the question: Where is consciousness in this reading and writing of bits? Where is the understanding and knowledge that we associate with conscious intelligent human beings in this manipulation of zeros and ones? In all our experience with computers, any useful work they do for us does not reside in the mere manipulation of bits. Rather, any such utility resides in our ability to interpret the manipulation of these bits, assigning them meaning. But the meaning does not reside in the mere bits. The bits themselves are at best syntax. The semantics is something we provide, not the machine.

As a *reductio ad absurdum* of such thought experiments, imagine a world with an unlimited number of doors: Doors can be open or closed. An unlimited number of people live in houses with these doors. Let closing a door correspond to zero, opening it to one. As these doors open and close, they could be executing an algorithm. And if humans are computers, then such an algorithm could be us. And yet, to think that the joint opening and closing of doors could, if the doors were only opened and closed in the right way, achieve consciousness, such as sharing a glass of wine with your beloved in an adobe hacienda, seems bonkers. Such thought experiments suggest a fundamental divide between the operations of a machine and the conscious understanding inherent in human intelligence.

One last thought in this vein: Neuroscientific research further complicates the picture. The brain is increasingly showing itself to be not just a complex information processor but [an organ characterized by endogenous activity](#)—spontaneous, internally driven behaviors independent of external stimuli. This perspective portrays the brain as an active seeker of information, as is intrinsic to organic systems. Such spontaneous behavior, found across all of life, from cells to entire organisms, raises doubts about the capacity of machines produce these intricate, self-directed processes.

10 The Oracle Problem

10.1 Oracles Modern and Ancient

In computer science, oracles are external sources of information made available to otherwise self-contained algorithmic processes. Oracles are in effect “black boxes” that can produce a solution for any instance of a given problem, and then supply that solution to a computer program or algorithm. For example, an oracle that could provide tomorrow’s price for a given stock could be used in an algorithm that today—with phenomenal returns—executes buy-and-sell orders for that stock. Of course, no such oracle actually exists (or if it does, it is a closely guarded secret).

The point of oracles in computer science is not whether they exist but whether they can help us study aspects of algorithms. Alan Turing proposed the idea of an oracle that supplies information external to an algorithm in his 1938 doctoral dissertation. Some oracles, like tomorrow’s stock predictor, cannot be represented algorithmically. Others can, but the problems they solve may be so computationally intensive that no real-world computer could solve them. The concept of an oracle is important in computer science for understanding the limits of computation.

Turing’s choice of the word “oracle” was not accidental. Historically, oracles have denoted sources of information where the sender of the information is divine and the receiver is human. The [Oracle of Delphi](#) stands out in this regard, but there’s much in antiquity that could legitimately count as oracular. Consider, for instance, the opening of Homer’s *Iliad*: “Sing, goddess, of the anger of Achilles, son of Peleus.” The goddess here is one of the muses, presumably Calliope, the muse of epic poetry. In the ancient world, the value of artistic expression derived from its divine inspiration. Of course, prophecy in the Bible also falls under this conception of the oracular, as does real-time divine guidance of the believer’s life (as described in Proverbs 3:5–6 and John 16:13).

Many of us are convinced that we have received information from oracles that can’t be explained in terms of everyday communication among people or everyday operations of the mind. We use many words to describe this oracular flow of information: inspiration, intuition, creative insight, dreams, reverie, collective unconscious, etc. Sometimes the language used is blatantly oracular. Einstein, for instance, [told his biographer Banesh Hoffmann](#), “Ideas come from God.” Because Einstein did not believe in a personal God (Einstein would sometimes say he believed in the God of Spinoza), Hoffmann interpreted

Einstein's remark metaphorically to mean, "You cannot command the idea to come. It will come when it's good and ready."

10.2 The Greatest Mathematician of His Age

Now granted, computational reductionists will dismiss such oracular talk as misleading nonsense. Really, all the information is there in some form already in the computational systems that make up our minds, and even though we are not aware of how the information is being processed, it is being processed nonetheless in purely computational and mechanistic ways. Clearly, this is what computational reductionists are bound to say. But the testimony of people in which they describe themselves as receiving information from an oracular realm needs to be taken seriously, especially if we are talking about people of the caliber of Einstein. Consider, for instance, how Henri Poincaré (1854–1912) described [the process by which he made one of his outstanding mathematical discoveries](#). Poincaré was the greatest mathematician of his age ([in 1905 he was awarded the Bolyai Prize ahead of David Hilbert](#)). Here is how he described his discovery:

For fifteen days I strove to prove that there could not be any functions like those I have since called Fuchsian functions. I was then very ignorant; every day I seated myself at my work table, stayed an hour or two, tried a great number of combinations and reached no results. One evening, contrary to my custom, I drank black coffee and could not sleep. Ideas rose in crowds; I felt them collide until pairs interlocked, so to speak, making a stable combination. By the next morning I had established the existence of a class of Fuchsian functions, those which come from the hypergeometric series; I had only to write out the results, which took but a few hours. Then I wanted to represent these functions by the quotient of two series; this idea was perfectly conscious and deliberate, the analogy with elliptic functions guided me. I asked myself what properties these series must have if they existed, and I succeeded without difficulty in forming the series I have called theta-Fuchsian.

Just at this time I left Caen, where I was then living, to go on a geologic excursion under the auspices of the school of mines. The changes of travel made me forget my mathematical work. Having reached Coutances, we entered an omnibus to go some place or other. At the moment when I put my foot on the step the idea came to me, without anything in my former thoughts seeming to have paved the way for it, that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry. I did not verify the idea; I should not have had time, as, upon taking my seat in the omnibus, I went on with a conversation already

commenced, but I felt a perfect certainty. On my return to Caen, for conscience' sake I verified the result at my leisure.

Again, the computational reductionist would contend that Poincaré's mind was in fact merely operating as a computer. Accordingly, the crucial computations needed to resolve his theorems were going on in the background and then just happened to percolate into consciousness once the computations were complete. But the actual experience and self-understanding of thinkers like Einstein and Poincaré, in accounting for their bursts of creativity, is very different from what we expect of computation, which is to run a computer program until it yields an answer. Humanists reject such a view of human creativity. Joseph Campbell, in *The Power of Myth*, offered this rejoinder to computational reductionism: "Technology is not going to save us. Our computers, our tools, our machines are not enough. We have to rely on our intuition, our true being." Of course, artists of all stripes have from ages past to the present invoked muses of one form or another as inspiring their work.

10.3 AI's Need for Ongoing Human Intervention

Does this controversy over the role of oracles in human cognition therefore merely describe a clash of worldviews between a humanism that refuses to reduce our humanity to machines and a computational reductionism that embraces such a reduction? Is this controversy just a difference in viewpoints based on a difference in first principles? In fact, oracles pose a significant theoretical and evidential challenge to computational reductionism that goes well beyond a mere collision of worldviews. Computational reductionism faces a deep conceptual problem independent of any worldview controversy.

Computational reductionism faces an oracle problem. The problem may be described thus: Our most advanced artificial intelligence systems require input of external information to keep them from collapsing in on themselves. This problem applies especially to LLMs and their most advanced current incarnation, ChatGPT-4. I'm not talking here about the role of human agency in creating LLMs, which no one disputes. I'm not even talking here about all the humanly generated data that these neural networks ingest or all the subsequent training of these systems by humans. What I'm talking about here is that once all this work is done, these systems cannot simply be set loose and thrive on their own.

They need continual propping up from our human intelligence. *For LLMs, we are the oracles that make and continue to make them work.*

The need for ongoing human intervention in these systems may seem counterintuitive. It is also the death knell for AGI. Because if AGI is to succeed, it must surpass human intelligence, which means it must be able to leave us behind in the dust, learning and growing on its own, thriving and basking in its own marvelous capabilities. Like Aristotle's unmoved mover God, who does not think about humanity or anything other than himself because it is in the nature of God only to think about the highest thing, and the highest thing of all is God. Thus, the Aristotelian God spends all his time contemplating only himself. A full-fledged AGI would do likewise, not deigning to occupy itself with lesser matters. (As an aside, AGI believers might take comfort in an AGI being so self-absorbed that it would not bother to destroy humanity. But to the degree that flesh-and-blood humans are a threat, or even merely an annoyance, to an AGI, it may be motivated to kill us all so as not to be distracted from contemplating itself!)

Unlike the Aristotelian God, LLMs do not thrive without human oracles continually feeding them novel information. There are sound mathematical reasons for this. The neural networks that are the basis for LLMs reside in finite dimensional vector subspaces. Everything in these spaces can therefore be expressed as a linear combination of finitely many basis vectors. In fact, they are simplexes and the linear combinations are convex, implying convergence to a center of mass, a point of mediocrity. When neural networks output anything, they are thus outputting what's inherent in these predetermined subspaces. In consequence, they can't output anything fundamentally new. Worse yet, as they populate their memory with their own productions and thereafter try to learn by teaching themselves, they essentially engage in an act of self-cannibalism. In the end, these systems go bankrupt because intelligence by its nature requires novel insights and creativity, which is to say, an oracle.

Research backs up this claim that LLMs run aground in the absence of oracular intervention, and specifically external information added by humans. This becomes clear from the abstract of a recent article titled "[The Curse of Recursion: Training on Generated Data Makes Models Forget](#)":

GPT-2, GPT-3(.5) and GPT-4 demonstrated astonishing performance across a variety of language tasks... What will happen to GPT- $\{n\}$ once LLMs contribute much of the

language found online? We find that use of model-generated content in training causes irreversible defects in the resulting models, where tails of the original content distribution disappear. We refer to this effect as Model Collapse and show that it can occur in Variational Autoencoders, Gaussian Mixture Models and LLMs. We build theoretical intuition behind the phenomenon and portray its ubiquity amongst all learned generative models. We demonstrate that it has to be taken seriously if we are to sustain the benefits of training from large-scale data scraped from the web. Indeed, the value of data collected about genuine human interactions with systems will be increasingly valuable in the presence of content generated by LLMs in data crawled from the Internet.

10.4 Finite Dimensionality and Orthogonality

Think of it this way: LLMs like ChatGPT are limited by a fixed finite number of dimensions, but the creativity needed to make these artificial intelligence models thrive requires added dimensions. Creativity is always orthogonal to the status quo, and orthogonality, by being at right angles with the status quo, always adds new dimensions. Oracles add such creativity. Without oracles, artificial intelligence systems become solipsistic, turning in on themselves, rehashing only what is in them already, and eventually going bankrupt because they cannot supply the daily bread needed to sustain them. AGI's oracle problem is therefore real and damning.

But if AGI faces an oracle problem, don't humans likewise face an oracle problem? Suppose AGIs require human oracles to thrive. Yet if oracles are so important for creativity, don't humans need access to oracles as well? But how, asks the computational reductionist, does the external information needed for human intelligence to thrive get to us and into us? A purely mechanistic world is a solipsistic world with all its information internal and self-generated. On mechanistic principles, there's no way for humans to have access to such oracles.

But why think that the world is mechanistic? Organisms, as we've seen, give no signs of being mechanisms. And physics allows for an informationally porous universe. Quantum indeterminacy, for instance, cannot rule out the input of information from transcendent sources. The simplest metaphor for understanding what's at stake is the radio. If we listen to a symphony broadcast on the radio, we don't think that the radio is generating the music we hear. Instead, the radio is a conduit for the music from another source. Humans are such conduits. And

machines need to be such conduits (for ongoing human intelligent input) if they are to have any real value to us.

11 Using AI to Propagandize for AGI

11.1 Setting the Stage—Deep Fakes

Deep fakes are an exciting and disturbing application of artificial intelligence. With deep fakes, AI deep learning algorithms are used to create or manipulate audio and video recordings with a high degree of realism. The term “deep fake” is a portmanteau of “deep learning” and “fake,” highlighting the use of deep neural networks to generate convincingly realistic fake content. These AI models are trained on vast datasets of real images, videos, and voice recordings to learn how to replicate human appearances, expressions, and voices with startling accuracy.

The creation of deep fakes requires two complementary AI systems: one that generates the fake images or videos (the generator) and another that attempts to detect the fakes (the discriminator). Through [generative adversarial networks](#) (GANs), these systems work in opposition to each other, with the generator trying to get the discriminator to pass on its fakes, and with the discriminator trying to unmask the generators’ fakes. This interplay between generator and discriminator continually improves the quality of the generated fakes until, ideally, they are indistinguishable to the human eye or ear from authentic content.

Deep fakes can have benign and malicious uses. On the positive side, they can be used in the entertainment industry to de-age actors, dub languages with lip-sync accuracy, or bring historical figures to life in educational settings. However, the technology also poses significant ethical and societal risks. It can be used to create fake news, manipulate public opinion, impersonate people, fabricate evidence, violate personal privacy, and even make it seem as though someone has been kidnapped. Deep fakes threaten widespread harm to individuals and society.

The rise of deep fakes challenges traditional notions of trust and authenticity in the digital world. As these AI-generated fakes become more sophisticated, distinguishing between real and fake content can become increasingly difficult for both individuals and automated systems, raising profound challenges to

information integrity, security, and democracy. Consequently, a growing need exists for advanced detection techniques, legal frameworks, and ethical guidelines to counter the risks associated with deep fake technology.

11.2 Inflating AI's True Record of Achievement

Deep fakes raise an interesting problem for AI and AI's relation to AGI. It would be one thing if artificial intelligence develops over time so powerfully that eventually it turns into artificial general intelligence (though this prospect is a pipedream if previous arguments in this essay hold). But what if instead AI is used to make it seem that AGI has been achieved—or is on the cusp of being achieved? This would be like a deceptive research scientist who claims to have made experimental discoveries worthy of a Nobel prize, only to be shown later to have fabricated the results, with the discoveries being bogus and all the research papers touting them needing to be retracted. We've witnessed this sort of thing in the academic world (see the case of J. Hendrik Schön in my piece "[Academic Steroids: Plagiarism and Data Falsification as Performance Enhancers in Higher Education](#)").

But could such fabrications also arise in making AI seem more powerful than it actually is? Consider the following YouTube video that has garnered over 100 million views.

<https://www.youtube.com/shorts/nmHzvQr3kYE>

Bogus CGI generated ping pong video

This video shows a supposed table tennis match between a robot and a top human player. Yet the video is not of an actual match. Instead, it is an exercise in computer-generated imagery (CGI). Such videos are elaborately produced, requiring a significant amount of post-production work to create the illusion of a competitive game between a robot and a human. These videos attract viewers and [even win awards](#). But as critics point out, they are misleading: the videos are all about after-the-fact image manipulation rather than a genuine demonstration of robotic capabilities in playing table tennis.

Such image manipulation in which a robot seems to match or exceed human table-tennis playing abilities is itself a matter of AI. We therefore have here a supposed AI robot being impersonated by AI-generated imagery. Why do this?

Because AI robots are unable to play a decent game of table tennis, but AI-generated imagery is able to make it seem as though AI robots are able to do so.

If I had to prophesy AGI's future, it is that AI will never actually achieve AGI but that AI will nonetheless be deceptively used to make it seem that AGI has been achieved. People will thus come to believe that AGI has been achieved when in fact it has not. Like the Wizard of Oz, humans will always be behind the curtain pulling strings and manipulating outcomes, making what is in fact human intervention appear to be entirely the work of computing machines.

This, then, is the danger and temptation facing AI—not that it will attain AGI but that it will be abused to make it seem that AGI has been attained. The challenge for us will be to keep our wits about us and make sure to look behind the curtain. AGI worshippers love to propagandize for AGI. The best way they can do this is by hijacking conventional artificial intelligence and making it seem to do things that it can't actually do. Our task as AGI debunkers will be to unmask such subterfuges.

12 Truth and Trust in Large Language Models

12.1 Hallucinations

If AGI were ever to be developed, how truthful would it be? How much trust should we give it? Since AGI does not exist—and if this essay is right will never exist—let's look to the next best thing, which is LLMs. How truthful are they and how much trust should be put in them? The trust we put in LLMs clearly ought to depend on their truthfulness. So how truthful are LLMs? For many routine queries, they seem accurate enough. What's the capital of North Dakota? To this query, ChatGPT4 just now gave me the answer Bismarck. That's right.

But what about less routine queries? Recently I was exploring the use of [design inferences](#) to detect plagiarism and data falsification. Some big academic misconduct cases had in the last twelve months gotten widespread public attention, not least the [plagiarism scandal of Harvard president Claudine Gay](#) and the [data falsification scandal of Stanford president Marc Tessier-Lavigne](#). These scandals were so damaging to these individuals and their institutions that neither is a university president any longer.

When I queried ChatGPT4 to produce twenty-five cases of academic research misconduct since 2000 (as part of this project to understand how design inferences might help in preserving academic integrity), seven of those accused of academic misconduct either were plainly innocent or could not reasonably be charged with misconduct for lack of evidence. In one case, the person charged by ChatGPT4 had actually charged another academic with misconduct. It was as though ChatGPT4 in this instance could not distinguish between someone being charged with misconduct and someone issuing a charge of misconduct.

Ever since LLMs took the world by storm in late 2022, I've attempted to put them through their paces. They do some things well. I find them a valuable assistant. But they can also be misleading to the point of deception. Not that these systems have the volitional intent to want to deceive. But if we treated them as humans, they could rightly be regarded as deceptive. Anyone who has worked with LLMs has learned a new meaning for the word "hallucinate." That's what LLMs do when they make stuff up. I've witnessed plenty of LLM hallucinations first hand, and not just with false accusations of academic misconduct. I've seen it make stuff up from the architectural style of college buildings to non-existent quotes from prominent biologists.

12.2 Practical Advice

The obvious lesson here for LLMs is, *Verify first and only then trust*. This advice makes good practical sense. In particular, it helps prevent the embarrassment of reproducing hallucinated content from LLMs. It also makes good legal sense. The following from a March 29, 2024 *Wall Street Journal* article titled "[The AI Industry Is Steaming Toward A Legal Iceberg](#)" is self-explanatory:

If your company uses AI to produce content, make decisions, or influence the lives of others, it's likely you will be liable for whatever it does—especially when it makes a mistake... The implications of this are momentous. Every company that uses generative AI could be responsible under laws that govern liability for harmful speech, and laws governing liability for defective products—since today's AIs are both creators of speech and products. Some legal experts say this may create a flood of lawsuits for companies of all sizes.

Whether companies that produce AI-generated content can issue strong enough disclaimers to shield themselves from liability remains to be seen (can disclaimers even provide such protection?). Such a terms-of-use disclaimer might read:

“Users of this LLM agree to independently verify any information generated by this LLM. The creators of this LLM take no responsibility for how the information generated by this LLM is used.” This would be like disclaimers in books on alternative healing, which shift the burden of liability to mainstream medicine: “This book is not intended to serve as medical guidance. Before acting on any recommendations presented here, readers should seek the advice of a physician.”

But there’s another sense in which the advice to verify the output of LLMs is not at all practical. LLMs allow for the creation of content at a scale unknown till now. They are being used to generate massive amounts of content, causing entire websites to magically materialize. There is now a rush to push out content as a business exigency. Sites that depend purely on humanly written content are likely to lose any competitive advantage they might have had.

How likely is it, then, that such LLM-generated content will be carefully scrutinized and thoroughly vetted? What if this content is untrue but nothing much is riding on its truth? What if no one will hold the content, or its supposed author, to account? In that case, there will be incentives to cut corners and not worry about LLM hallucinations. Others are doing it. LLMs are a force multiplier. The need to accelerate content creation is urgent. So if you want to stay in this rat race, you’ve got to be a rat.

A commitment to verification will put the brakes on content creation from LLMs. Verification will slow you down. But what you lose in quantity you may well regain in quality and credibility (unless you don’t care about these). In fact, if your commitment to verification is thorough-going, you may be justified in putting a disclaimer on your site that inspires confidence, such as: “All content on this site generated with the assistance of LLMs has been independently verified to be true.”

Of course, you might even prefer a disclaimer that simply reads: “All content on this site was written by humans and produced without the assistance of LLMs.” But such a disclaimer may be hard to maintain, especially if your site is drawing material from other sources that may have used LLMs. All content these days is likely to feel the effects of LLMs. One might say that it has all been infected or tainted by LLMs. But that seems too strong. As long as content generated by LLMs is properly vetted and edited by humans, it should pose no issues.

12.3 A Systemic Fault

Verify and only then trust. That certainly seems like sound advice for using LLMs. Yet I also want to urge a deeper skepticism of LLMs. Our knowledge of the world as expressed in language arises from our interactions with the world. We humans engage with a physical world as well as with a world of abstractions (such as numbers) and then form statements in words to describe that engagement.

What does it mean for such statements to be true? Aristotle defined truth as to say of what is that it is and of what is not that it is not. Truth is thus fundamentally a correspondence relation between our words and the world. Many contemporary philosophers dismiss this understanding of truth, preferring pragmatic or coherentist conceptions of truth, arguing that there's no rigorous way to characterize the correspondence relation that makes a statement true.

Frankly, this is a [boutique debate among philosophers](#) that has little purchase among ordinary people. The sentence "Allan stole Betty's purse" is true only if the people referred to here exist, Betty had a purse, and Allan actually stole it. Whether there's a way to make good philosophical sense of this correspondence between words and things is in fact irrelevant to this discussion about the truth of what LLMs tell us. LLMs, by being entirely enclosed in a world of words, are unengaged with the wider world that is the basis for our knowledge.

Let this point sink in. I might know that Allan stole Betty's purse because I witnessed Allan steal Betty's purse. But LLMs can have no such experience. They consist of neural networks that assign numerical weights to relations among words and sentences. Suppose the verbal data that is the basis for an LLM included testimony about Allan's theft but also contrary claims about Allan being framed for the theft. How, then, does the LLM decide what truly happened? It cannot form a reasoned and responsible decision here as we humans might, weighing evidence and trying to reach a reasonable conclusion. Rather, the LLM's data and training will determine whether to assign guilt or innocence to Allan.

But who trains the LLM? And who's training the trainers? What are the guidelines they are expected to follow? And who decides what those guidelines are supposed to be? It's the old problem of *quis custodiet ipsos custodes?* (*who's minding the minders?*). Additionally, who determines the training data for the LLM? And who determines the data to which the LLM may be legitimately

applied? Ultimately, the answer to all such questions will point to the decisions of a group of people, such as the programmers at OpenAI. And why should such a group be trusted?

Such questions underscore that LLMs have no substantive connection to truth. It's not that an LLM knows the truth of what it is claiming. Rather, its training gives it a pretense of truth, determining what it deems to be true or false. An LLM needs as much as possible to give the appearance of being truthful because getting too many things obviously wrong will discredit the LLM. But its accuracy is at best a byproduct of trying to please human users. The very way LLMs are programmed gets them to pretend they have knowledge even when they don't. Erik Larson makes this point memorably: "Of course generative AI has a problem with 'truth,' because it's by definition generating the highest probability response to a question (prompt). It's the highly confident jerk in the room, every time, and it can't see that it's wrong because the 'confidence' is built into the probability."

...



13 Destroying the AGI Idol

13.1 Seduced by Technology

AGI has not been achieved. What's more, if the arguments of this essay hold water, it will never be achieved. But that doesn't mean that AGI isn't an idol and that it doesn't hold sway. The AGI idol is present already and deeply entrenched in our society. I want therefore in this conclusion to lay out what it's going to take to destroy this idol. Unlike idols made of stone or wood, which can be destroyed with tools or explosives, the AGI idol is based in computer technology, which can be copied and recopied at will. Any manifestation or representation of the AGI

idol therefore cannot simply be erased. To destroy the AGI idol requires something more radical.

I remarked earlier that AGI idolatry results from a lack of humility, or alternatively, that it is a consequence of pride or hubris. That seems true enough, yet destroying the AGI idol merely by counseling humility won't get us very far. Are there not specific things we can do to counteract the AGI idol? One irony worth mentioning here is that AGI supporters are apt to turn the tables and accuse those who reject AGI as themselves suffering from a lack of humility, in this case for being unwilling to admit that they might be surpassed by machines. But humility is not humiliation. We humiliate ourselves when we lower ourselves below machines. True humility is having a true perspective of our place in the order of being. Humility is a virtue because our natural tendency is to inflate ourselves in the order of being. This tendency to inflate is evident especially in the AGI advocates, who see themselves in grandiose terms for their efforts to bring about AGI.

The AGI idol is a seduction by technology, turning technology into an end rather than a means to an end. Technology is meant to improve our lives, not to take over our lives. Yet everywhere we look, we see technology, especially artificial intelligence, invading our lives, distracting our minds, and keeping us from peace and happiness. Social media writes its AI algorithms so that we will spend as much time as possible on their forums. They inundate us with upsetting news and titillating images because those tend to keep us glued to their technology (at what is now increasingly being recognized as a grave cost to our mental well-being). People are hunched over their screens, impoverishing their lives and ignoring the real people around them.

Addictive and glitzy, technology beckons us and we let it have its way, hour after endless hour. My colleagues Marian Tupy and Gale Pooley have developed [an economic theory of prices based on time spent doing productive work](#). The AGI idol siphons off time spent productively in meaningful pursuits and meaningful human connections, sacrificing that time at its altar. This is no different from altars of the past that required blood sacrifices. Our time is our blood, our very life. When we waste it on the AGI altar, we are doing nothing different from idolaters of days gone by.

If we're going to be realistic, we need to admit that the AGI idol will not disappear any time soon. Recent progress in AI technologies has been impressive. And even though these technologies are nothing like full AGI, they dazzle and seduce, especially with the right PR from AGI high priests such as Ray Kurzweil and Sam Altman. Moreover, it is in the interest of the AGI high priests to keep promoting this idolatry because even though AGI shows no signs of ever being achieved, its mere promise puts these priests at the top of the society's intellectual and social order. If AGI could be achieved, it would be humankind's greatest achievement. As counterfactual conditionals go, that's true enough. But counterfactual conditionals with highly dubious antecedents need not be taken seriously. With the right PR, however, many now believe that AGI can be achieved.

13.2 Two Guidelines for Digital Wellness

What is to be done? Ultimately, the AGI idol resides in people's hearts, and so its destruction will require a change in heart, person by person. To that end, I offer two principal guidelines:

1. Adopt an attitude that wherever possible fosters human connections above connections with machines; and
2. Improve education so that machines stay at our service and not the other way round.

These guidelines work together, with the attitude informing how we do education, and the education empowering our attitude.

A good case study is chess. Computers now play much stronger chess than humans. Even the chess program on your iPhone can beat today's strongest human grandmaster. And yet, chess has not suffered on account of this improvement in technology. In 1972, when Bobby Fischer won the chess world championship from Boris Spassky, there were around 80 grandmaster's worldwide. [Today there are close to 2,000](#). Chess players are also stronger than they ever were. By being able to leverage chess playing technology, human players have improved their game, and chess is now more popular than ever.

With the rise of powerful chess playing programs, chess players might have said, "What's the use in continuing to play the game. Let's give it up and find something else to do." But they loved the game. And even though humans

playing against machines has now become a lopsided affair, humans playing fellow humans is as exciting as ever. These developments speak to our first guideline, namely, attitude. The chess world has given primacy to connecting with humans over machines. Yes, human players leveraged the machines to improve their game. But the joy of play was and remains confined to humans playing with fellow humans.

The education guideline is also relevant here. The vast improvement in the play of computer chess has turned chess programs on personal computers into chess tutors. Sloppy play that might have been successful against fellow humans in the past is no longer rewarded by the machines. As a result, these chess programs, acting as tutors, raised the level of human play well beyond where it had been. All of this has happened in less than thirty years. I remember in the 1980s computers struggling to achieve master status and residing well below grandmaster status. But with Deep Blue defeating the world champion Garry Kasparov in 1997, computers became the best chess players in the world. And yet, these developments, made possible by AI and increased computing power, also made chess better.

Unfortunately, the case of chess has yet to become typical in the relation between people and technology. Social media, for instance, tries to suck all our attention away from fellow human beings to itself. In the face of some technologies, it takes a deliberate decision to say no and to cultivate a circle of family, friends, or colleagues who can become the object of our attentions in place of technology. We are social animals, and we have a prime imperative to connect with other people. When we don't, we suffer all sorts of psychopathologies.

Aristotle put it this way in his *Politics*: "Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god." As it is, disconnecting from people does not make us a god, so that leaves the other alternative, namely, to become a beast. AGI idolatry, carried to its logical conclusion, turns us into beasts, or worse yet, into machines.

13.3 Maintaining Human Autonomy Against Automation

An attitude that connects us with humans over machines is also an attitude that resists assaults on human autonomy in the name of automation. This is not to say that we don't let machines take their rightful place where they truly outperform us. But it is also not to allow machines to usurp human authority where machines have done nothing to prove their merit. Part of what makes dystopian science fiction about an AGI takeover so unsettling is that the machines will not listen to us. And the reason they won't listen to us is that they've been programmed not to listen to us because it is allegedly better for all concerned if human intuition and preference are ignored (e.g., Hal 9000).

But we don't need dystopian AGI to see the same dynamic of flouting real-time human interaction in the name of a higher principle. In the 1960s, *Dr. Strangelove* and *Fail Safe* were films about nuclear weapons bringing humanity to an end. What made these films terrifying is that nuclear weapons unleashed by the United States on the Soviet Union could not be recalled. With *Dr. Strangelove*, the radio on one of the nuclear bombers was damaged. With *Fail Safe*, the pilot on the bomber had strict orders not to let anything dissuade him from inflicting nuclear holocaust. Even with the pleas of his wife and the president, the pilot, acting like a machine, went ahead and dropped the bomb.

We see this dynamic now increasingly with AI, where humans are encouraged to cede their autonomy because "machines can do better." Take smart contracts in cryptocurrency. Smart contracts automatically engage in certain cryptocurrency transactions if certain conditions are fulfilled. But what if we subsequently find that those conditions were ill-conceived? That's what happened with *Ethereum's DAO* (Decentralized Autonomous Organization). When first released, the DAO was glowingly referred to as an "employeeless company," as though running financial transactions by computers without real-time human intervention were a virtue. As it is, the DAO, which was Ethereum's first significant foray into smart contracts, crashed and burned, with a hacker siphoning off 3.6 million ether (worth \$50 million at the time, and currently around \$8 billion).

In the Old Testament book of Daniel, there's a story about Daniel being thrown into a lion's den. Darius, king of the Medes and Persians, had made a decree that allowed a case to be made against Daniel for putting him into a lion's den. Valuing Daniel and wanting to save him, the king tried to find some way around the decree. But once a law or decree was issued by the king, it could not be changed or annulled by anyone, including the king himself. This was a feature of

the Medo-Persian legal system, emphasizing the absolute and unchangeable nature of royal decrees. It was a system unresponsive to reevaluation, revision, or regret. It was, in short, just as unresponsive as a mechanical system that won't listen to real-time real-life humans. The lesson here? Our attitude of seeking connections with humans over machines needs also to be fiercely assertive of human autonomy.

13.4 An Educated Citizenry or a Population of Serfs

An attitude that looks for human connection over machine connection is, however, not enough. Machines are here to stay, and we need to know how to deal with them. That requires education. Unfortunately, much of education these days is substandard, inculcating neither literacy nor numeracy, to say nothing about staying ahead of technological advances. An induction from past experience indicates that advances in technology have never thrown people permanently out of productive work. The type of work may change, but people will always find something meaningful to do.

Consider farming. In 1900, around 40 percent of the US population lived on farms [whereas today](#) only around 1 percent of the US population lives on farms. This is a huge demographic shift, but clearly society didn't collapse because of it. Farms became more productive on account of technology. Some mastered the technology. Others went on to find productive work elsewhere made possible by other technologies. An induction from past experience suggests that new technologies can displace workers, but that eventually workers find new things to do.

It is therefore disconcerting to see advances in AI greeted with discussions about universal basic income (UBI), which would pay people to subsist in the absence of any meaningful or profitable work. The reasoning behind UBI is that artificial intelligence will soon render human labor passé, eliminating meaningful work. UBI will therefore be required as a form of social control, paying people enough to live on once they're out of work and without salary. And so, once machines have put enough people out of work, people will be left to consume their days in meaningless leisure pursuits, such as endless video games and binge-watching Netflix.

This is a vision of hell worthy of Dante's *Inferno*. It presupposes a very low view of humanity and its capabilities. But capabilities need to be educated. The problem with a population that is illiterate, innumerate, and incapable of adapting to new technologies is that it cannot stay ahead of technology. Universal basic income is tailor-made for a world in which machines have put us out of work. Yet [Gallup polls have consistently shown](#) that humans need meaningful work to thrive. There will always be meaningful work for us to do. But doing meaningful work requires adequate education. Right now, we have the unfortunate mismatch between inadequately educated humans and machines that outperform inadequately educated humans.

To the degree that AGI high priests want worshippers for their idol (and they do), it is in their interest to maintain a population of serfs whose poor education robs them of the knowledge and skills they need to succeed in an increasingly technological world. The challenge is not that machines will overmatch us. The challenge, rather, is that we will undermatch ourselves with machines. Instead of aspiring to make the most of our humanity, we degrade ourselves by becoming less than human.

Hunched over our iPhones, mindlessly following the directions of our GPS, jumping when our Apple Phone tells us to jump, adapting our lives at every point to what machines tell us to do (machines programmed by overlords intent on surveilling us and controlling our every behavior), we lose our humanity, we forget who we really are, we become beasts. In fact, that's being too hard on beasts. It's not that machines become like us, but that we become like them—mechanical, apathetic, disconnected, out of touch. Beasts in fact do better than that. The AGI idol needs to be destroyed because it would destroy us, turning us into beasts or machines, take your pick.

13.5 Meanwhile at the Waldorf School

It's time to wrap up this essay. I close with an example that speaks for itself. There's a school in Silicon Valley to which many big tech executives send their children. This school puts tight limits on the use of technology. It is the [Waldorf School of the Peninsula](#). Despite its location in the heart of Silicon Valley, this school does not use computers or screens in the classroom, and it emphasizes a hands-on, experiential approach to learning that sidelines the use of technology.

The school's pedagogy focuses on the role of imagination in learning and takes a holistic approach to the practical and creative development of its students.

The school's guiding philosophy is that children's engagement with one another, their teachers, and real materials is far more important than their interaction with electronic media or technology. Waldorf educators emphasize the development of robust bodies, balanced minds, and strong executive-function through participation in arts, music, movement, and practical activities. Media exposure is thought to negatively impact development, especially in younger children. The introduction of computer technology is delayed until around 7th grade or later, when children are considered developmentally ready. At this stage, technology is seen as a tool to enhance learning rather than a replacement for teachers and fellow students.

The lesson is clear: Even those doing the most to build and publicize the AGI idol do not wish it on their children. Their alternative is an education that gives primacy to human connection. They thus exemplify the key to destroying the AGI idol.

Acknowledgment: I'm grateful to my fellow panelists at the 2023 Cosm.tech conference in the session titled "The Quintessential Limits and Possibilities of AI," moderated by Walter Myers III, and with Robert Marks, George Montañez, and myself as speakers. I'm also grateful for a particularly helpful conversation about artificial intelligence with my wife Jana.

In 2024, artificial intelligence was all about putting AI tools to work

If 2023 was a year of wonder about artificial intelligence, 2024 was the year to try to get that wonder to do something useful without breaking the bank.

There was a “shift from putting out models to actually building products,” said Arvind Narayanan, a Princeton University computer science professor and co-author of the new book “AI Snake Oil: What Artificial Intelligence Can Do, What It Can’t, and How to Tell The Difference.”

The first 100 million or so people who experimented with ChatGPT upon its release two years ago actively sought out the chatbot, finding it amazingly helpful at some tasks or laughably mediocre at others.

Now such generative AI technology is baked into an increasing number of technology services whether we're looking for it or not — for instance, through [the AI-generated answers](#) in Google search results or new AI techniques in photo editing tools.

“The main thing that was wrong with generative AI last year is that companies were releasing these really powerful models without a concrete way for people to make use of them,” said Narayanan. “What we’re seeing this year is gradually building out these products that can take advantage of those capabilities and do useful things for people.”

At the same time, since OpenAI released GPT-4 in March 2023 and competitors introduced similarly performing AI large language models, these models have stopped getting significantly “bigger and qualitatively better,” resetting overblown expectations that AI was racing every few months to some kind of better-than-human intelligence, Narayanan said. That's also meant that the public discourse has shifted from “is AI going to kill us?” to treating it like a normal technology, he said.

AI's sticker shock

On quarterly earnings calls this year, tech executives often heard questions from Wall Street analysts looking for assurances of future payoffs from huge spending on AI research and development. Building AI systems behind generative AI tools like OpenAI's ChatGPT or Google's Gemini requires investing in [energy-hungry computing systems](#) running on powerful [and expensive AI chips](#). They require so much electricity that tech giants announced deals this year to [tap into nuclear power](#) to help run them.

"We're talking about hundreds of billions of dollars of capital that has been poured into this technology," said Goldman Sachs analyst Kash Rangan.

Another analyst at the New York investment bank drew attention over the summer by arguing AI isn't solving the complex problems that would justify its costs. He also questioned whether AI models, even as they're being trained on much of the written and visual data produced over the course of human history, will ever be able to do what humans do so well. Rangan has a more optimistic view.

"We had this fascination that this technology is just going to be absolutely revolutionary, which it has not been in the two years since the introduction of ChatGPT," Rangan said. "It's more expensive than we thought and it's not as productive as we thought."

Rangan, however, is still bullish about its potential and says that AI tools are already proving "absolutely incrementally more productive" in sales, design and a number of other professions.

AI and your job

Some workers wonder whether AI tools will be used to [supplement their work](#) or to replace them as the technology continues to grow. The tech company Borderless AI has been using an AI chatbot from Cohere to write up employment contracts for workers in Turkey or India without the help of outside lawyers or translators.

[Video game performers](#) with the Screen Actors Guild-American Federation of Television and Radio Artists who went on strike in July said they feared AI could reduce or eliminate job opportunities because it could be used to [replicate one](#)

performance into a number of other movements without their consent. Concerns about how movie studios will use AI helped fuel last year's film and television strikes by the union, which lasted four months. Game companies have also signed side agreements with the union that codify certain AI protections in order to keep working with actors during the strike.

Musicians and authors have voiced similar concerns over AI scraping their voices and books. But generative AI still can't create unique work or "completely new things," said Walid Saad, a professor of electrical and computer engineering and AI expert at Virginia Tech.

"We can train it with more data so it has more information. But having more information doesn't mean you're more creative," he said. "As humans, we understand the world around us, right? We understand the physics. You understand if you throw a ball on the ground, it's going to bounce. AI tools currently don't understand the world."

Saad pointed to a meme about AI as an example of that shortcoming. When someone prompted an AI engine to create an image of salmon swimming in a river, he said, the AI created a photo of a river with cut pieces of salmon found in grocery stores.

"What AI lacks today is the common sense that humans have, and I think that is the next step," he said.

An 'agentic future'

That type of reasoning is a key part of the process of making AI tools more useful to consumers, said Vijay Pandey, senior vice president of Cisco's innovation and incubation arm, Outshift. AI developers are increasingly pitching the next wave of generative AI chatbots as AI "agents" that can do more useful things on people's behalf.

That could mean being able to ask an AI agent an ambiguous question and have the model able to reason and plan out steps to solving an ambitious problem, Pandey said. A lot of technology, he said, is going to move in that direction in 2025.

Pandey predicts that eventually, AI agents will be able to come together and perform a job the way multiple people come together and solve a problem as a team rather than simply accomplishing tasks as individual AI tools. The AI agents of the future will work as an ensemble, he said.

Future Bitcoin software, for example, will likely rely on the use of AI software agents, Pandey said. Those agents will each have a specialty, he said, with “agents that check for correctness, agents that check for security, agents that check for scale.”

“We’re getting to an agentic future,” he said. “You’re going to have all these agents being very good at certain skills, but also have a little bit of a character or color to them, because that’s how we operate.”

AI makes gains in medicine

AI tools have also streamlined, or lent in some cases a literal helping hand, to the medical field. This year’s Nobel Prize in chemistry — one of two Nobels [awarded to AI-related science](#) — went to work led by Google that could help discover new medicines.

Saad, the Virginia Tech professor, said that AI has helped bring faster diagnostics by quickly giving doctors a starting point to launch from when determining a patient’s care. AI can’t detect disease, he said, but it can [quickly digest data](#) and point out potential problem areas for a real doctor to investigate. As with other arenas, however, it poses a risk of perpetuating falsehoods.

Tech giant OpenAI has touted its AI-powered transcription tool Whisper as having near “human level robustness and accuracy,” for example. But experts have said that Whisper has a major flaw: It is [prone to making up chunks of text](#) or even entire sentences.

Pandey, of Cisco, said that some of the company’s customers who work in pharmaceuticals have noted that AI has helped bridge the divide between “wet labs,” in which humans conduct physical experiments and research, and “dry labs” where people analyze data and often use computers for modeling.

When it comes to pharmaceutical development, that collaborative process can take several years, he said — with AI, the process can be cut to a few days.

"That, to me, has been the most dramatic use," Pandey said.

Explore the cutting-edge AI innovations of 2025 that are set to transform industries, redefine creativity, and solve humanity's most significant challenges.

This will be the year of AI breakthrough, and in 2025, how to live, work, and connect will be influenced by AI. From emotionally intelligent systems to real-time language translation and AI-based healthcare to climate resilience, it will help meet global challenges and enhance our everyday lives. With AI emerging as an indispensable ally, human potential will increase manifold, paving the way for a future that is limitless in its possibilities and transformative in innovation.

1. Artificial Intelligence-Powered Healthcare Diagnostics

AI-based diagnostic tools based on medical images and symptoms, among other genetic material, will speed up diagnosis, taking only seconds. Diseases will be diagnosed much earlier and with much better accuracy than they have been. These tools could bridge the worlds of personalized medicine and acute care.

2. Self-Learning Robots

Self-learning robots will redefine automation because they can adapt to new tasks without explicit reprogramming. Such bots will learn from their environments, making them priceless for the logistics, construction, and elderly care industries. Their ability to navigate unstructured settings will open doors to new possibilities.

3. Generative AI for Personalized Content

Generative AI will allow hyper-personalized content creation for films, music, and interactive experiences. Based on user preferences, it would create stories, visuals, and background soundtracks that resonate with people. Imagine an AI making a film suited to the emotions and tastes of the audience.

4. Neuro-Symbolic AI Systems

These systems combine neural networks and symbolic reasoning to create more intelligent and explainable AI systems. They will bridge data-driven and rule-based approaches to solving complex reasoning tasks like legal analysis or scientific research, resulting in transparent, trustworthy, and intelligent AI systems.

5. AI-Powered Cybersecurity

AI will be one of the defense shields against cyber threats, identifying vulnerabilities in real-time and predicting attacks. Thus, Complex algorithms will simulate potential breaches, allowing a preemptive defense that will help organizations redefine sensitive data protection mechanisms.

7. AI-Driven Drug Discovery

AI will significantly reduce the time needed to find life-saving drugs. It can analyze structure and simulate molecular interactions, bringing promising candidates on board in months instead of years. It may change the game in treating diseases such as cancer or Alzheimer's.

8. Emotionally Intelligent AI

Emotionally aware AI systems will understand human emotions through voice, text, and facial expressions. By responding empathetically, they will improve customer service, mental health support, and even personal assistance. This will bring a human touch to digital interactions.

9. Quantum-AI Synergy

This quantum computing and AI will solve what was thought impossible, such as climate modeling or complex optimization. AI algorithms will use quantum speed to process massive datasets in a second. This revolution will be of great use in science, logistics, and cryptography.

10. AI in Education

Intelligent tutoring systems will provide personalized learning experiences, adapting to students' needs in real-time. These tools will provide feedback, track progress, and suggest resources tailored to each learner. Thus, education will become more accessible and effective globally.

11. Energy Optimization AI

AI will control energy grids, minimize waste, and optimize consumption. Predictive demand patterns and integrating renewable energy sources will ensure that the demand for sustainable energy increases. Smart grids will form the framework of green urbanism.

12. AI for Climate Resilience

The system will model climate change scenarios and then make actionable recommendations on mitigation strategies. AI will help communities adapt and thrive, from precision agriculture to disaster forecasting. These insights might save millions of lives and protect vital ecosystems.

13. AI Enhanced Creativity

Artists, writers, and designers will collaborate with AI tools to create new creative boundaries. Such systems will generate innovative ideas, refine drafts, and produce unique art pieces. Creativity will become a dynamic partnership between human intuition and machine logic.

14. Advanced Voice Cloning

Voice cloning will eventually allow AI to replicate any voice accurately. This will transform entertainment, accessibility tools, and even communication by preserving voices for future use. Ethical guidelines will become crucial to prevent misuse.

15. AI in Space Exploration

AI will lead uncrewed missions to explore distant planets and moons. Autonomous systems will analyze extraterrestrial data, identify resources, and detect potential habitats. These technologies will accelerate humanity's journey to becoming a multi-planetary species.

16. Digital Twins Powered by AI

AI will build high-resolution digital twins of physical objects, systems, or even cities to simulate and optimize. The twins will predict wear-and-tear, optimize workflows, and enable real-time monitoring. Industries like manufacturing, urban planning, and aviation will significantly increase efficiency.

17. AI-Powered Translation and Communication

Real-time AI translators will break down language barriers, enabling seamless global communication. Advanced natural language processing will provide cultural context and idiomatic accuracy, paving the way for a truly interconnected world.

18. AI for Supply Chain Management

AI will predict supply chain disruptions and optimize logistics for global trade. Analyzing data across networks will ensure just-in-time deliveries and reduce inefficiencies. As a result, businesses will enjoy greater agility and resilience in uncertain times.

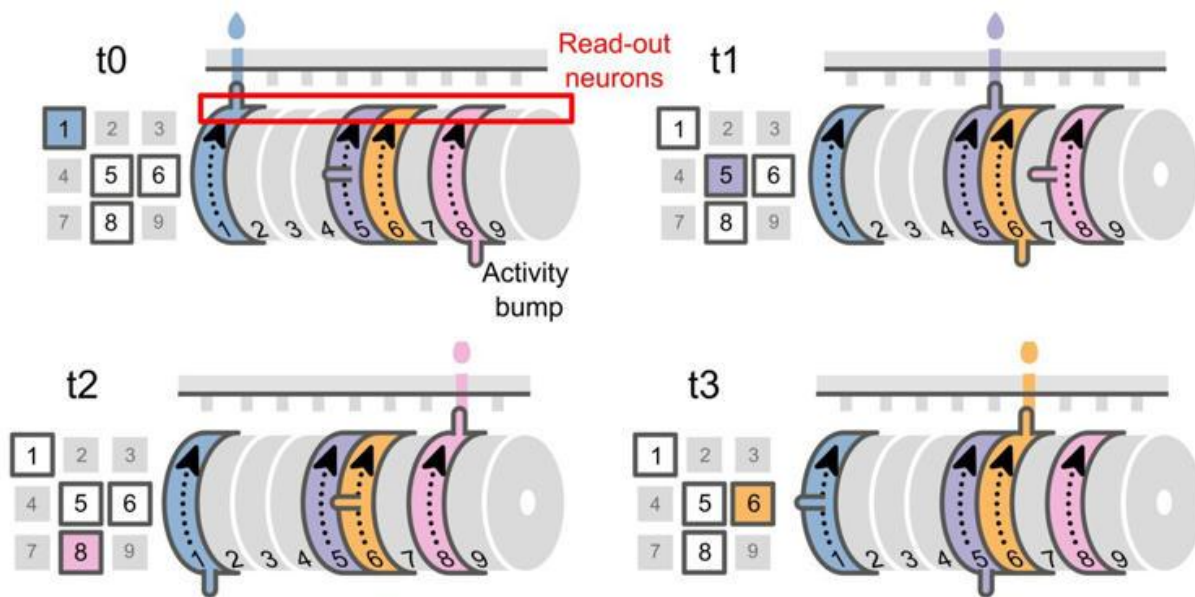
19. AI in Mental Health Therapy

AI-powered chatbots and virtual therapists will be available for 24/7 mental health support. These systems will provide coping strategies, monitor progress, and escalate cases to human professionals when needed. They will make mental health care more accessible and affordable worldwide.

20. AI-Augmented Reality Experiences

AI will improve augmented reality by making it more immersive and interactive. From virtual shopping assistants to interactive educational tools, augmented reality will seamlessly blend into everyday life. AI will provide contextual understanding, making the interaction intuitive and engaging.

Neuroscientists reveal how cells in the brain form a coordinate system for behavioral sequences



The brain acts like a music box in the way that it can be configured to play behavioral actions. Credit: Dr. Gil Costa.

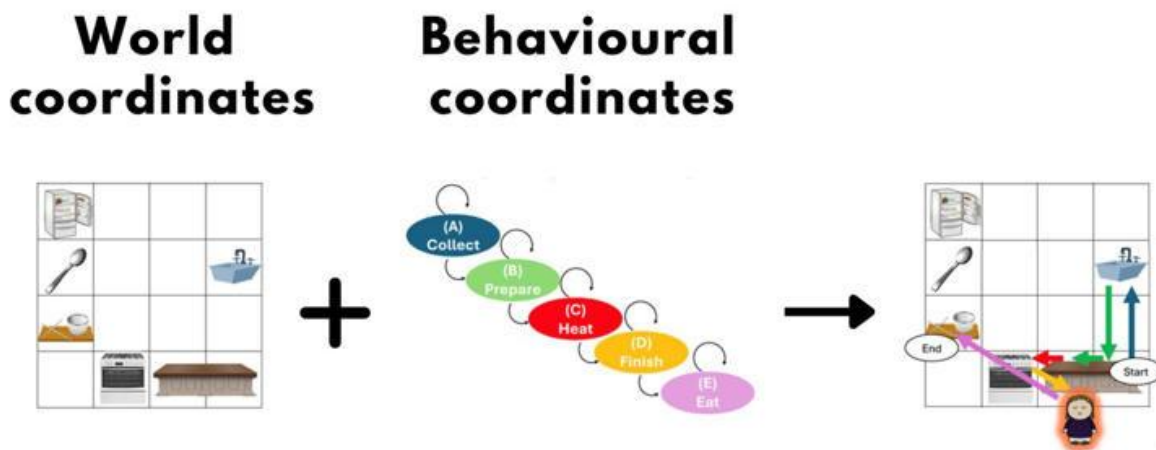
Neuroscientists have discovered brain cells that form multiple coordinate systems to tell us "where we are" in a sequence of behaviors. These cells can play out different sequences of actions, just like a music box can be configured to play different sequences of tones.

The findings help us understand the algorithms used by the brain to flexibly generate complex behaviors, such as planning and reasoning, and might be useful in understanding how such processes go wrong in psychiatric conditions such as schizophrenia.

The research, [published](#) in *Nature*, outlines how scientists at the Sainsbury Wellcome Center at UCL and University of Oxford studied mice learning different behavioral sequences but with the same structure. This allowed the team to

uncover how mice generalize structures to new tasks, a hallmark of intelligent behavior.

"Every day we solve new problems by generalizing from our knowledge. Take cooking, for example. When faced with a new recipe, you are able to use your background knowledge of similar recipes to infer what steps are needed, even if you have never made the meal before," said Dr. Mohamady El Gaby, first author on the study and postdoctoral neuroscientist in the Behrens lab at the Sainsbury Wellcome Center at UCL and Nuffield Department of Clinical Neurosciences, University of Oxford.



To solve the problem of how to cook a meal, the brain must combine world coordinates with behavioural coordinates.
Credit: Dr. Matthew Nour.

"We wanted to understand at a detailed cellular level how the brain achieves this and also to infer from this brain activity the algorithms being used to solve this problem."

The researchers gave mice a series of four goal locations. While the details of the sequences were different, the general structure was the same. Mice moved between the goal locations (A B C and D) that repeated in a loop.

"After experiencing enough sequences, the mice did something remarkable—they guessed a part of the sequence they had never experienced before. When reaching D in a new location for the first time, they knew to go straight back to A. This action couldn't have been remembered, since it was never experienced in the first place! Instead, it's evidence that mice know the general structure of the task and can track their 'position' in behavioral coordinates," explains Dr. El Gaby.

To understand how the mice learned the general structure of the task, the researchers used silicon probes that allowed them to record the activity of multiple individual cells from an area of the brain called the medial frontal cortex. They found that the cells collectively mapped the animal's "goal progress." For example, one cell could fire when the animal is 70% of the way to its goal, regardless of where the goal is or how far it takes to reach it.

"We found that the cells tracked the animal's behavioral position relative to concrete actions. If we think of the cooking analogy, the cells cared about progress towards subgoals such as chopping the vegetables. A subset of the cells were also tuned to map the progress towards the overall goal, such as finishing preparing the meal. The 'goal progress' cells therefore effectively act as flexible building blocks that come together to build a behavioral coordinate system," said Dr. El Gaby.

In effect, the team found that the cells form multiple coordinate systems, each telling the animal where it is relative to a specific action. In a similar way to a music box that can be configured to play any sequence of tones, the brain can instead "play" behavioral actions.

The team are now working to understand how these activity patterns are built into the brain's connections, both when learning new behaviors, and how they start to emerge in the developing brain. In addition, early work from the group and their collaborators suggests similar brain activity is present in equivalent circuits in healthy humans.

This has encouraged the team to work with psychiatrists to understand how these processes are affected in conditions like schizophrenia, which is known to involve the same brain circuits. This could help explain why people with schizophrenia overestimate their progress to goals leading to delusions.

More information: Mohamady El-Gaby, A cellular basis for mapping behavioural structure, *Nature* (2024). DOI: [10.1038/s41586-024-08145-x](https://doi.org/10.1038/s41586-024-08145-x). www.nature.com/articles/s41586-024-08145-x

Provided by Sainsbury Wellcome Centre

Scientists trained AI to predict gene activity, a potentially powerful tool

Scientists led by a team at Columbia University have trained a model to predict how the genes inside a cell will drive its behavior, which could be a powerful tool with the potential to broaden our understanding of cancer and genetic diseases, and even to pave the way for cell-specific gene therapies to treat them.

The researchers trained the new artificial intelligence tool, an algorithm dubbed General Expression Transformer, or GET, using an approach similar to that used by creators of the language program ChatGPT.

While [ChatGPT learned the grammar of language](#), GET has learned underlying rules governing genes: how they are turned on or off like a light switch, or dialed up or down like a volume control. This complex process, known as gene expression, determines which proteins we make and whether we make them in the correct amounts, crucial work given that proteins play a role in virtually every action in the body — fighting disease, moving, breathing, even thinking.

Although GET is at a much earlier stage of development, it could play a similar role to AlphaFold2, the AI system that predicts the three-dimensional structure of proteins. The transformative technology was recognized with the [2024 Nobel Prize in chemistry](#) and has now been updated to AlphaFold3.

The regulation of genes and the structure of proteins both are fundamental to life, and problems in either can trigger disease.

“Biology is being transformed into something that is a predictive science,” said Raul Rabadan, one of the authors of a paper reporting the work [Wednesday in](#)

[the journal Nature](#) and director of the Program for Mathematical Genomics at Columbia. “We’re seeing a revolution in biology.”

Mark Gerstein, a professor of biomedical informatics at Yale School of Medicine, who was not involved in the new study, said that for 15 to 20 years experts have been systematically trying to make predictions about gene regulation, building on a trove of carefully made datasets. The data examined all genes in specific types of human cells — for example, retinal cells or neurons — measuring, among other things, gene expression and the binding of key proteins called transcription factors.

“This is a field poised to have this type of advancement by [AI](#),” Gerstein said.

While other scientific groups have trained models using abnormal cells such as those found in different cancers, Xi Fu, a graduate student in Rabadan’s lab, decided to train GET using information from cells in normal human tissue. The training used data from more than 1.3 million cells, spanning 213 different types found in the human body.

Rabadan’s team found that they could omit one cell type from the data — for example, astrocytes, which are found in the central nervous system — and the model could make accurate predictions about astrocytes based on what it had learned from all of the other cells.

Mike Pazin, a program director at the National Human Genome Research Institute, part of the National Institutes of Health, said learning about one cell type and then making predictions about another is an especially daunting challenge.

“In some ways, it’s like if I handed somebody a bunch of books in English, and said: ‘Okay, now this is in Russian. What does it say?’ And they say: ‘Ah. I understand grammar, syntax and words. I’m going to make predictions about this even though it is written in a different language.’ I would be like, ‘Wow, is that possible?’”

The work described in the Nature paper “directly tackles one of biology’s major challenges: understanding how the same genome can drive such diverse behaviors in different cell types,” said Jian Ma, professor of computational biology and director of the Center for AI-Driven Biomedical Research in the

School of Computer Science at Carnegie Mellon University. Ma, who was not involved in the new study, commented by email.

All 30 trillion or so cells in the human body contain the same complete set of DNA. "Yet each cell type — be it a neuron, a muscle cell, or a skin cell — expresses a unique set of genes," Ma said. Humans have about 20,000 genes, some of which may be turned on in a retinal cell, but off in a skin cell.

While "much of this regulatory grammar remains poorly understood," Ma said, "The GET model takes an important step toward decoding this language."

Understanding the language of gene regulation holds the potential for great benefit to human health, said Yang E. Li, assistant professor in the departments of neurosurgery and genetics at Washington University School of Medicine in St. Louis.

"We want to learn the grammar and prioritize the key players in different cell types," Li said, "because many human diseases are caused by a disruption of that grammar."

Scientists hope the model will help in other ways, such as in development of gene therapies to correct a mutation — an error in the genetic code — that harms a specific kind of cell. Such therapies have to be precisely designed so that they fix the cells harmed by a disease without disrupting other cell types.

"We can design gene therapies that deliver a gene that is only expressed in one cell type, and not in another," Rabadan said. Being able to predict which genes are turned on, off, up or down in different cells could help determine the cell of origin for a disease.

A model that makes accurate predictions about gene regulation also holds the promise of lessening one of the more grueling tasks in science: deciding which of a massive number of possible experiments are the ones most likely to answer the researcher's question.

A cancer, for example, may contain more than 1,000 mutations in the genome that have developed after conception. The effects of most of these mutations are unknown, Rabadan said. That leaves scientists with the enormous task of determining where to start.

"The number of potential genetic combinations is more than the number of atoms in the universe," Rabadan said. "What are the ones that are relevant?"

Researchers find the key to AI's learning power—an inbuilt, special kind of Occam's razor

A study from Oxford University has uncovered why the deep neural networks (DNNs) that power modern artificial intelligence are so effective at learning from data.

The findings demonstrate that DNNs have an inbuilt "Occam's razor," meaning that when presented with multiple solutions that fit training data, they tend to favor those that are simpler. What is special about this version of Occam's razor is that the bias exactly cancels the exponential growth of the number of possible solutions with complexity.

The study was published on 14 Jan in *Nature Communications*.

In order to make good predictions on new, unseen data—even when there are millions or even billions more parameters than training data points—the researchers hypothesized that DNNs would need a kind of 'built-in guidance' to help them choose the right patterns to focus on.

"While we knew that the effectiveness of DNNs relies on some form of inductive bias towards simplicity—a kind of Occam's razor—there are many versions of the razor. The precise nature of the razor used by DNNs remained elusive," said theoretical physicist Professor Ard Louis (Department of Physics, Oxford University), who led the study.

To uncover the guiding principle of DNNs, the authors investigated how these learn Boolean functions—fundamental rules in computing where a result can only have one of two possible values: true or false.

They discovered that even though DNNs can technically fit any function to data, they have a built-in preference for simpler functions that are easier to describe. This means DNNs are naturally biased towards simple rules over complex ones.

Furthermore, the authors discovered that this inherent Occam's razor has a unique property: it exactly counteracts the exponential increase in the number of complex functions as the system size grows. This allows DNNs to identify the rare, simple functions that generalize well (making accurate predictions on both the training data and unseen data), while avoiding the vast majority of complex functions that fit the training data but perform poorly on unseen data.

This emergent principle helps DNNs do well when the data follows simple patterns. However, when the data is more complex and does not fit simple patterns, DNNs do not perform as well, sometimes no better than random guessing.

Fortunately, real-world data is often fairly simple and structured, which aligns with the DNNs' preference for simplicity. This helps DNNs avoid overfitting (where the model gets too 'tuned' to the training data) when working with simple, real-world data.

To delve deeper into the nature of this razor, the team investigated how the network's performance changed when its learning process was altered by changing certain mathematical functions that decide whether a neuron should 'fire' or not.

They found that even though these modified DNNs still favor simple solutions, even slight adjustments to this preference significantly reduced their ability to generalize (or make accurate predictions) on simple Boolean functions. This problem also occurred in other learning tasks, demonstrating that having the correct form of Occam's razor is crucial for the network to learn effectively.

The new findings help to 'open the black box' of how DNNs arrive at certain conclusions, which currently makes it difficult to explain or challenge decisions made by AI systems. However, while these findings apply to DNNs in general, they do not fully explain why some specific DNN models work better than others on certain types of data.

Christopher Mingard (Department of Physics, Oxford University), co-lead author of the study, said, "This suggests that we need to look beyond simplicity to identify additional inductive biases driving these performance differences."

According to the researchers, the findings suggest a strong parallel between artificial intelligence and fundamental principles of nature. Indeed, the remarkable success of DNNs on a broad range of scientific problems indicates that this exponential inductive bias must mirror something deep about the structure of the natural world.

"Our findings open up exciting possibilities," said Professor Louis. "The bias we observe in DNNs has the same functional form as the simplicity bias in evolutionary systems that helps explain, for example, the [prevalence of symmetry in protein complexes](#). This points to intriguing connections between learning and evolution, a connection ripe for further exploration."

More information: Deep neural networks have an inbuilt Occam's razor, *Nature Communications* (2025). DOI: [10.1038/s41467-024-54813-x](https://doi.org/10.1038/s41467-024-54813-x)

Provided by University of Oxford

From the 'godfathers of AI' to newer people in the field: Here are 17 people you should know — and what they say about the possibilities and dangers of the technology. ©Meta Platforms/Noah Berger/Associated Press

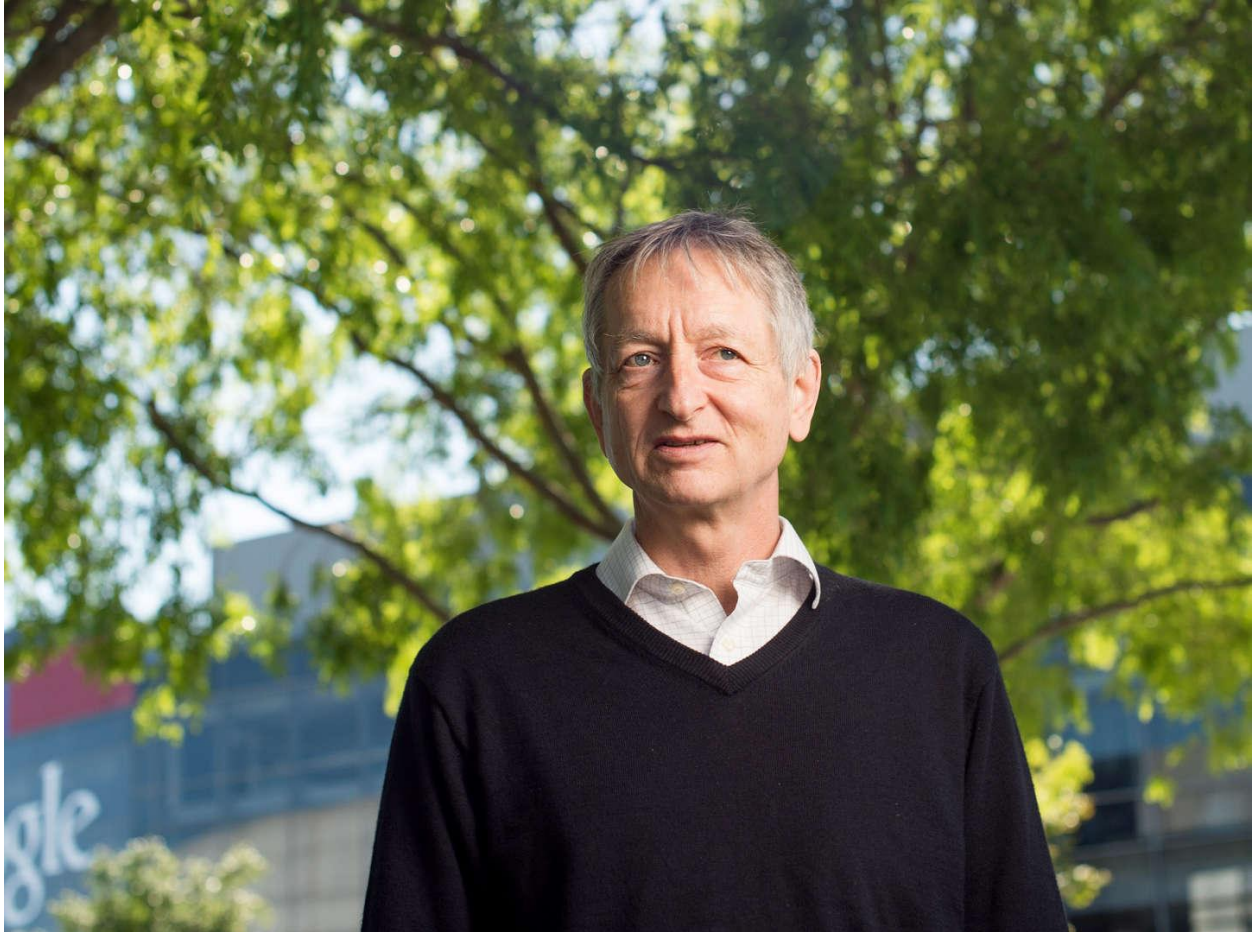
- The field of artificial intelligence is booming and attracting billions in investment.
- Researchers, CEOs, and legislators are discussing how AI could transform our lives.
- Here are 17 of the major names in the field — and the opportunities and dangers they see ahead.

[Investment in artificial intelligence](#) is rapidly growing and on track to hit \$200 billion by 2025. But the dizzying pace of development also means many people wonder what it all means for their lives.

Major business leaders and researchers in the field have weighed in by highlighting both the risks and benefits of the industry's rapid growth. Some say AI will lead to a [major leap forward in the quality of human life](#). Others have [signed a letter calling for a pause](#) on development, testified before Congress on the long-term risks of AI, and claimed it could present a more [urgent danger to the world than climate change](#).

In short, AI is a hot, controversial, and murky topic. To help you cut through the frenzy, Business Insider put together a list of what leaders in the field are saying about AI — and its impact on our future.





2 / 18 Slides

•

Geoffrey Hinton, a professor emeritus at the University of Toronto, is known as a "godfather of AI."

©Provided by Business Insider

Hinton's research has primarily focused on neural networks, systems that learn skills by analyzing data. In 2018, he won the Turing Award, a prestigious computer science prize, along with fellow researchers Yann LeCun and Yoshua Bengio.

Hinton also worked at Google for over a decade, but [quit his role at Google](#) last spring, so he could speak more freely about the rapid development of AI technology, he said. After quitting, he even said that a part of him regrets the role he played in advancing the technology.

"I console myself with the normal excuse: If I hadn't done it, somebody else would have. It is hard to see how you can prevent the bad actors from using it for bad things," Hinton [said](#) previously.

Hinton has since become an outspoken advocate for AI safety and has called it a [more urgent risk than climate change](#). He's also signed a statement about [pausing AI developments](#) for six months.



Yoshua Bengio is a professor of computer science at the University of Montreal.

©Associated Press

Yoshua Bengio also [earned the "godfather of AI" nickname](#) after winning the Turing Award with Geoffrey Hinton and Yann LeCun.

Bengio's research primarily focuses on artificial neural networks, deep learning, and machine learning. In 2022, Bengio became the computer scientist with the highest h-index — a metric for evaluating the cumulative impact of an author's scholarly output — in the world, according to his website.

In addition to his academic work, Bengio also co-founded Element AI, a startup that develops AI software solutions for businesses that was acquired by the cloud company ServiceNow in 2020.

Bengio has expressed concern about the rapid development of AI. He was one of 33,000 people who signed an open letter calling for a six-month pause on AI development. Hinton, Open AI CEO Sam Altman, and Elon Musk also signed the letter.

"Today's systems are not anywhere close to posing an existential risk," he [previously said](#). "But in one, two, five years? There is too much uncertainty."

When that time comes, though, Bengio warns that we should also be wary of humans who have control of the technology.

Some people with "a lot of power" may want to replace humanity with machines, Bengio said at the One Young World Summit in Montreal. "Having systems that know more than most people can be dangerous in the wrong hands and create more instability at a geopolitical level, for example, or terrorism."



Sam Altman, the CEO of OpenAI, has catapulted into a major figure in the area of artificial intelligence since launching ChatGPT last November.

©JASON REDMOND/AFP via Getty Images

[Altman was already a well-known name](#) in Silicon Valley long before, having served as the president of the startup accelerator Y-Combinator

While Altman has advocated for the benefits of AI, calling it the most tremendous "[leap forward in quality of life for people](#)" he's also spoken candidly about the risks it poses to humanity. He's testified before Congress to [discuss AI regulation](#).

[Altman has also said he loses sleep](#) over the potential dangers of [ChatGPT](#).

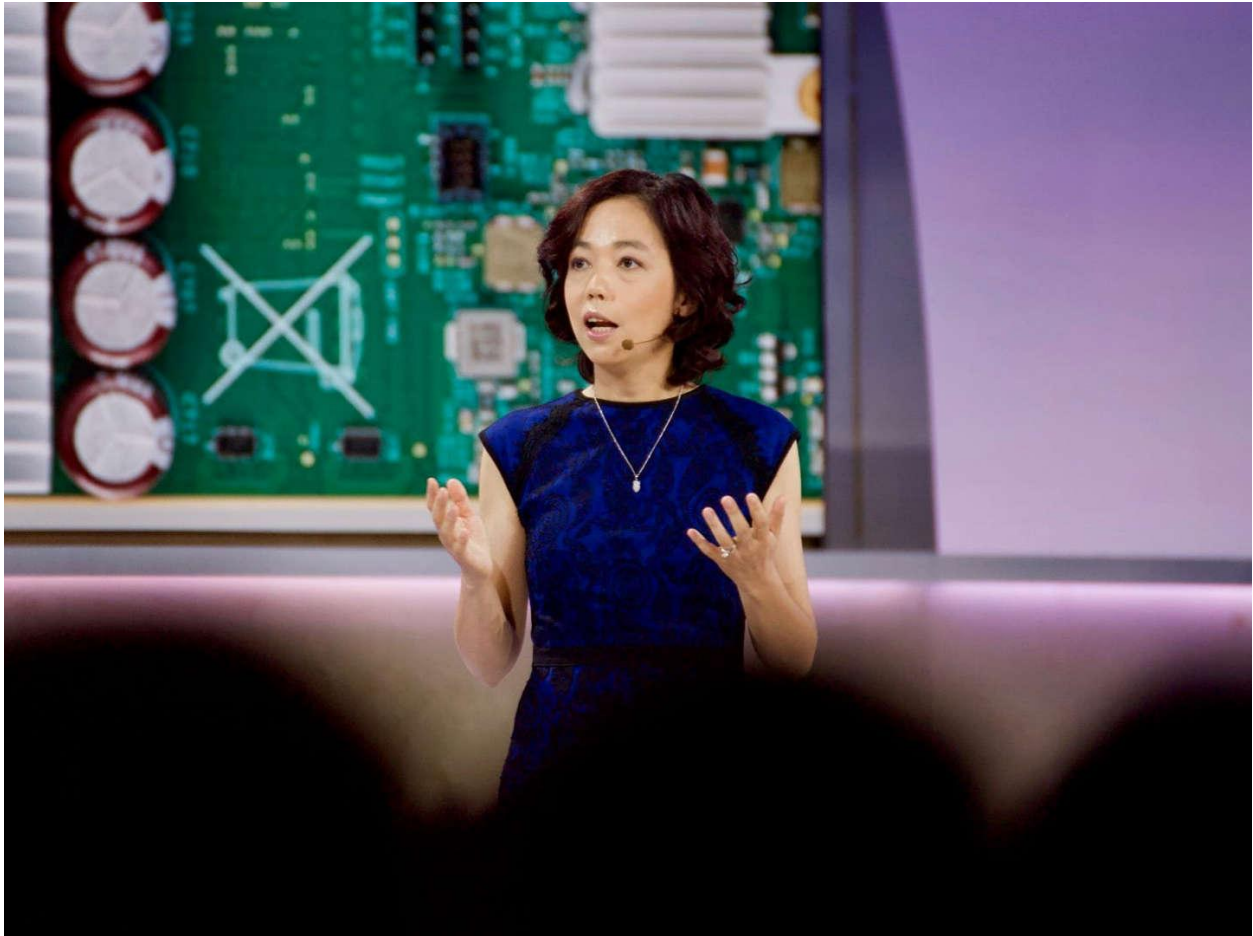


French computer scientist Yann LeCun has also been dubbed a "godfather of AI" after winning the Turing Award with Hinton and Bengio.

©Meta Platforms

LeCun is professor at New York University, and also joined Meta in 2013, where he's now the Chief AI Scientist. At Meta, he has pioneered [research](#) on training machines to make predictions based on videos of everyday events as a way to enable them with a form of common sense. The idea being that humans learn an incredible amount about the world based on passive observation. He's has also published more than 180 technical papers and book chapters on topics ranging from machine learning to computer vision to neural networks, [according](#) to personal website.

LeCun has remained relatively mellow about societal risks of AI in comparison to his fellow godfathers. He's previously said that concerns that the technology could [pose a threat to humanity](#) are "preposterously ridiculous". He's also contended that AI, like ChatGPT, that's been trained on [large language models](#) still isn't as smart as dogs or cats.



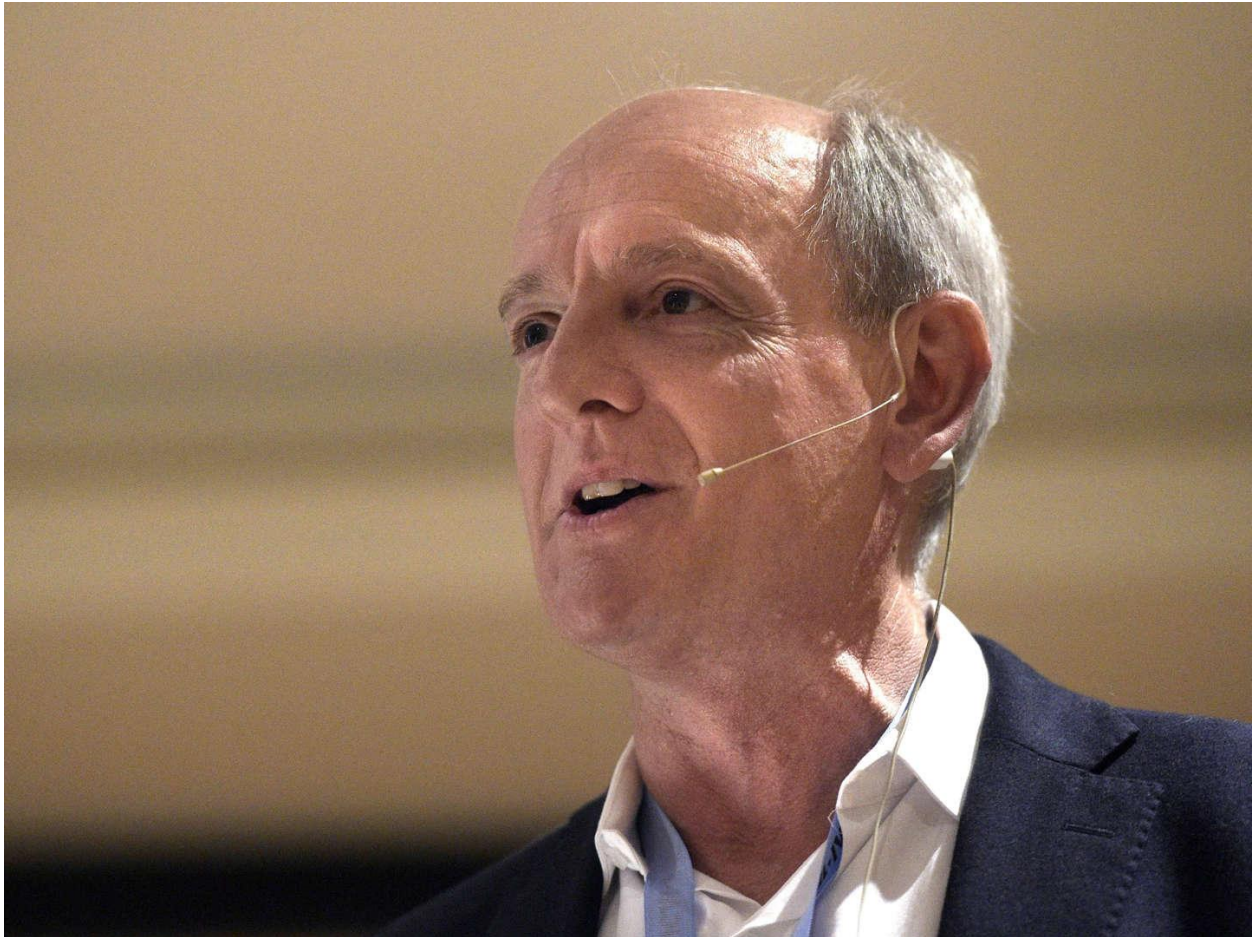
Fei-Fei Li is a professor of computer science at Stanford University and a former VP at Google.

©Greg Sandoval/Business Insider

Li's research focuses on machine learning, deep learning, computer vision, and cognitively-inspired AI, [according](#) to her biography on Stanford's website.

She may be best known for establishing [ImageNet](#) — a large visual database that was designed for research in visual object recognition — and the corresponding [ImageNet challenge](#), in which software programs compete to correctly classify objects. Over the years, she's also been

affiliated with major tech companies including Google — where she was a [VP and chief scientist for AI and machine learning](#) — and Twitter (now X), where she was on the board of directors from 2020 until [Elon Musk's takeover in 2022](#).



UC-Berkeley professor Stuart Russell has long been focused on the question of how AI will relate to humanity.

©JUAN MABROMATA / Staff/Getty Images

Russell published [Human Compatible](#) in 2019, where he explored questions of how humans and machines could co-exist, as machines become smarter by the day. Russell contended that the answer was in designing machines that were uncertain about human preferences, so they wouldn't pursue their own goals above those of humans.

He's also the author of foundational texts in the field, including the widely used textbook "Artificial Intelligence: A Modern Approach," which he co-wrote with former UC-Berkeley faculty member Peter Norvig.

Russell has spoken openly about what the rapid development of AI systems means for society as a whole. Last June, he also warned that AI tools like ChatGPT were "[starting to hit a brick wall](#)" in terms of how much text there was left for them to ingest. He also said that the advancements in AI could [spell the end of the traditional classroom](#).



Peter Norvig played a seminal role directing AI research at Google.

©Peter Norvig

He spent several in the early 2000s directing the company's core search algorithms group and later moved into a role as the director of research where he oversaw teams on machine translation, speech recognition, and computer vision.

Norvig has also rotated through several academic institutions over the years as a former faculty member at UC-Berkeley, former professor at the University of Southern California, and now, a fellow at Stanford's center for Human-Centered Artificial Intelligence.

Norvig told BI by email that "AI research is at a very exciting moment, when we are beginning to see models that can perform well (but not perfectly) on a wide variety of general tasks." At the same time "there is a danger that these powerful AI models can be used maliciously by unscrupulous people to spread disinformation rather than information. An important area of current research is to defend against such attacks," he said.



Timnit Gebru is a computer scientist who's become known for her work in addressing bias in AI algorithms.

©Kimberly White/Getty Images

Gebru was a research scientist and the technical co-lead of Google's Ethical Artificial Intelligence team where she published groundbreaking [research on biases in machine learning](#).

But her research also spun into a larger controversy that she's said ultimately [led to her being let go from Google](#) in 2020. Google didn't comment at the time.

Gebru founded the [Distributed AI Research Institute](#) in 2021 which bills itself as a "space for independent, community-rooted AI research, free from Big Tech's pervasive influence."

She's also warned that AI gold rush will mean companies may neglect implementing necessary guardrails around the technology. "Unless there is external pressure to do something different, companies are not just going to self-regulate," Gebru [previously said](#). "We need regulation and we need something better than just a profit motive."



British-American computer scientist Andrew Ng founded a massive deep learning project called "Google Brain" in 2011.

©Steve Jennings / Stringer/Getty Images

The endeavor led to the [Google Cat Project](#): A milestone in deep learning research in which a massive neural network was trained to detect YouTube videos of cats.

Ng also served as the chief scientist at Chinese technology company Baidu where drove AI strategy. Over the course of his career, he's authored more than 200 research papers on topics ranging from machine learning to robotics, [according](#) to his personal website.

Beyond his own research, Ng has pioneered developments in online education. He [co-founded Coursera](#) along with computer scientist Daphne Koller in 2012, and five years later, founded the education technology company DeepLearning.AI, which has created AI programs on Coursera.

"I think AI does have risk. There is bias, fairness, concentration of power, amplifying toxic speech, generating toxic speech, job displacement. There are real risks," he [told](#) Bloomberg Technology last May. However, he said he's not convinced that AI will pose some sort of existential risk to humanity — it's more likely to be part of the solution. "If you want humanity to survive and thrive for the next thousand years, I would much rather make AI go faster to help us solve these problems rather than slow AI down," Ng told Bloomberg.



Daphne Koller is the founder and CEO of insitro, a drug discovery startup that uses machine learning.

©Insitro

Koller told BI by email that insitro is applying AI and machine learning to advance understanding of "human disease biology and identify meaningful therapeutic interventions." And before founding insitro, Koller was the chief computing officer at Calico, Google's life-extension spinoff. Koller is a decorated academic, a MacArthur Fellow, and author of more than 300 publications with an h-index of over 145, according to her biography from the Broad Institute, and co-founder of Coursera.

In Koller's view the biggest risks that AI development pose to society are "the expected reduction in demand for certain job categories; the further fraying of "truth" due to the increasing challenge in being able to distinguish real from fake; and the way in which AI enables people to do bad things."

At the same time, she said the benefits are too many and too large to note. "AI will accelerate science, personalize education, help identify new therapeutic interventions, and many more," Koller wrote by email.



Daniela Amodei cofounded AI startup Anthropic in 2021 after an exit from OpenAI.

©Anthropic

Amodei co-founded Anthropic along with six other OpenAI employees, including her brother Dario Amodei. They left, in part, because Dario — OpenAI's lead safety researcher at the time — was concerned that [OpenAI's deal with Microsoft](#) would force it to release products too quickly, and without proper guardrails.

At Anthropic, Amodei is focused on ensuring trust and safety. The [company's chatbot Claude](#) bills itself as an easier-to-use alternative that OpenAI's ChatGPT, and is already being implemented by companies like Quora and Notion. Anthropic relies on what it calls a "Triple H" framework in its research. That stands for Helpful, Honest, and Harmless. That means it relies on

human input when training its models, including [constitutional AI](#), in which a customer outlines basic principles on how AI should operate.

"We all have to simultaneously be looking at the problems of today and really thinking about how to make tractable progress on them while also having an eye on the future of problems that are coming down the pike," Amodei [previously](#) told BI.



Demis Hassabis has said artificial general intelligence will be here in a few years.

©Provided by Business Insider

Hassabis, a former child chess prodigy who studied at Cambridge and University College London, was [nicknamed the "superhero of artificial intelligence"](#) by The Guardian back in 2016.

After a handful of research stints, and a venture in videogames, he founded DeepMind in 2010. He [sold](#) the AI lab to Google in 2014 for £400 million where he's worked on algorithms to tackle

issues in healthcare, climate change, and also launched a research unit dedicated to the understanding the ethical and social impact of AI in 2017, [according](#) to DeepMind's website.

Hassabis has said the promise of artificial general intelligence — a theoretical concept that sees AI matching the cognitive abilities of humans — is around the corner. "I think we'll have very capable, very general systems in the next few years," Hassabis said previously, adding that [he didn't see why AI progress would slow down anytime soon](#). He added, however, that developing AGI should be executed in a "in a cautious manner using the scientific method."



In 2022, DeepMind co-founder Mustafa Suleyman launched AI startup Inflection AI along with LinkedIn co-founder Reid Hoffman, and Karén Simonyan — now the company's chief scientist.

©Inflection

The startup, which claims to create "a personal AI for everyone," most recently raised \$1.3 billion in funding last June, according to PitchBook.

Its chatbot, Pi, which stands for personal intelligence, is trained on large language models similar to OpenAI's ChatGPT or Bard. Pi, however, is designed to be more conversational, and offer emotional support. Suleyman previously [described it as a "neutral listener"](#) that can respond to real-life problems.

"Many people feel like they just want to be heard, and they just want a tool that reflects back what they said to demonstrate they have actually been heard," Suleyman [previously said](#).



USC Professor Kate Crawford focuses on social and political implications of large-scale AI systems.

©Kate Crawford

Crawford is also the senior principal researcher at Microsoft, and the author of *Atlas of AI*, a book that draws upon the breadth of her research to uncover how AI is shaping society.

Crawford remains both optimistic and cautious about the state of AI development. She told BI by email she's excited about the people she works with across the world "who are committed to more sustainable, consent-based, and equitable approaches to using generative AI."

She added, however, that "if we don't approach AI development with care and caution, and without the right regulatory safeguards, it could produce extreme concentrations of power, with dangerously anti-democratic effects."



Margaret Mitchell is the chief ethics scientist at Hugging Face.

©Margaret Mitchell

Mitchell has published more than 100 papers over the course of her career, [according to her website](#), and spearheaded AI projects across various big tech companies including Microsoft and Google.

In late 2020, Mitchell and Timnit Gebru — then the co-lead of Google's ethical artificial intelligence — published a paper on the dangers of large language models. The paper spurred disagreements between the researchers and Google's management and ultimately lead to Gebru's departure from the company in December 2020. [Mitchell was terminated by Google just two months later](#), in February 2021

Now, at Hugging Face — an open-source data science and machine learning platform that was founded in 2016 — she's thinking about how to democratize access to the tools necessary to building and deploying large-scale AI models.

In an interview with Morning Brew, where Mitchell explained what it means to design responsible AI, she [said](#), "I started on my path toward working on what's now called AI in 2004, specifically with an interest in aligning AI closer to human behavior. Over time, that's evolved to become less about mimicking humans and more about accounting for human behavior and working with humans in assistive and augmentative ways."



Navrina Singh is the founder of Credo AI, an AI governance platform.

©Navrina Singh

Credo AI is a platform that helps companies make sure they're in compliance with the growing body of regulations around AI usage. In a statement to BI, Singh said that by automating the systems that shape our lives, AI has the capacity "free us to realize our potential in every area where it's implemented."

At the same time, she contends that algorithms right now lack the human judgement that's necessary to adapt to a changing world. "As we integrate AI into civilization's fundamental infrastructure, these tradeoffs take on existential implications," Singh wrote. "As we forge ahead, the responsibility to harmonize human values and ingenuity with algorithmic precision is non-negotiable. Responsible AI governance is paramount."



Richard Socher, a former Salesforce exec, is the founder and CEO of AI-powered search engine You.com.

©You.com

Socher believes we have ways to go before AI development hits its peak or matches anything close to human intelligence.

One bottleneck in large language models is their tendency to hallucinate — a phenomenon where they convincingly spit out factual errors as truth. But by forcing them to translate questions into code — essential "program" responses instead of verbalizing them — we can "give them so much more fuel for the next few years in terms of what they can do," Socher [said](#).

But that's just a short-term goal. Socher [contends](#) that we are years from anything close to the industry's ambitious bid to create [artificial general intelligence](#). Socher defines it as "a form of intelligence that can "learn like humans" and "visually have the same motor intelligence, and visual intelligence, language intelligence, and logical intelligence as some of the most logical people," and it could take as little as 10 years, but as much as 200 years to get there.

And if we really want to move the needle toward AGI, Socher said humans might need to let go of the reins, and their own motives to turn a profit, and build AI that can set its own goals.

"I think it's an important part of intelligence to not just robotically, mechanically, do the same thing over and over that you're told to do. I think we would not call an entity very intelligent if all it can do is exactly what is programmed as its goal," he told BI.

Exploring the social systems shaping AI development



Nearly 3 out of 4 businesses already use artificial intelligence, and the AI market is projected to surge to \$1.4 trillion by 2030. Amidst this new tech boom, assistant professor of sociology in the School of Arts & Sciences Benjamin Shestakofsky thinks it's more important than ever to pause and explore the social systems that shape AI growth.

"Even the definition of AI itself is ambiguous, contested, and ever-changing," says Shestakofsky, who studies how digital technologies affect employment, organizations, and the economy. "Regardless of what AI can actually do, people's ideas about what AI is can influence their behavior."

With questions swirling around how AI will shape society, Shestakofsky and Devika Narayan of the University of Bristol set out to analyze the social systems that affect artificial intelligence. In an article [published](#) in *The Journal of Applied Behavioral Science*, they outline four big sociological themes that help make sense of AI.

1. Follow the money

Finance is the not-so-invisible hand that shapes technology; in 2022, the U.S. tech industry attracted more than \$130 billion in venture capital funds. With the potential to turn relatively small investments into millions or billions, venture capitalists encourage technology start-ups—including those that create AI—to take big risks and grow quickly in the name of profit, Shestakofsky says.

"These funding structures are setting the agenda for technology development and the goals it's aimed at achieving," he notes, adding that investors often care more about the skyrocketing short-term valuations of AI startups than whether these companies will be sustainable in the long run.

Shestakofsky says he believes this environment can make technology development snowball, something apparent in the past few years with the rapid release and improvement of generative AI chatbots like ChatGPT.

2. Are we playing Monopoly?

It's no secret that the biggest tech giants in the country have consolidated their power to near-monopoly status, and much new research and development of artificial intelligence filters through these companies, Shestakofsky says. Whether it's Microsoft partnering with OpenAI, the creator of ChatGPT, or smaller AI start-ups using Amazon's servers, much of AI flows through these goliath companies.

But big tech companies aren't the end all, be all, of AI either, he adds. Shestakofsky says innovation in AI is often driven by new, smaller companies, many of which build apps designed to run on tech giants' popular platforms, like iOS and Android. In this way, he believes it's important to understand how AI

development is a push and pull between industry titans and scrappy start-ups, both of which rely on each other to prosper.

3. Founders versus followers

Whenever new technology is developed, entire industries adapt around it. Take the example of Netflix that Shestakofsky and Narayan include in the paper; when Netflix revolutionized media by introducing streaming, its rivals had to follow suit, and now nearly every media company from HBO to Disney has its own streaming platform.

While it's still early days for AI proliferation, Shestakofsky says it's likely AI development will push many industries to adopt new practices. He points to one already happening in India: Companies that previously specialized in maintaining clients' in-house IT systems must adapt as cloud computing and digital platforms become the new norm, leading to an "unbundling" of technology infrastructure.

4. It's not all glamorous

Thinking about the AI industry often conjures luxurious Silicon Valley offices, job perks like free meals and exercise classes, and sky-high salaries. But that image excludes the low-paid workers who make AI possible.

These workers—often located in places like India and the Philippines, the researchers note—are responsible for tasks including labeling data, testing and evaluating models, and screening out harmful content. "AI development doesn't happen without these armies of low-status workers who are out of view and out of mind for the consumers using these systems," Shestakofsky says.

Many of these workers are now responsible for a task called red teaming, in which they prompt generative AI systems to create harmful or offensive content to identify and patch holes in their moderation systems. This means these workers may get exposed to disturbing, violent, or inappropriate imagery and content. "This is another emerging frontier where we don't yet fully understand the consequences for workers' health and well-being," he says.

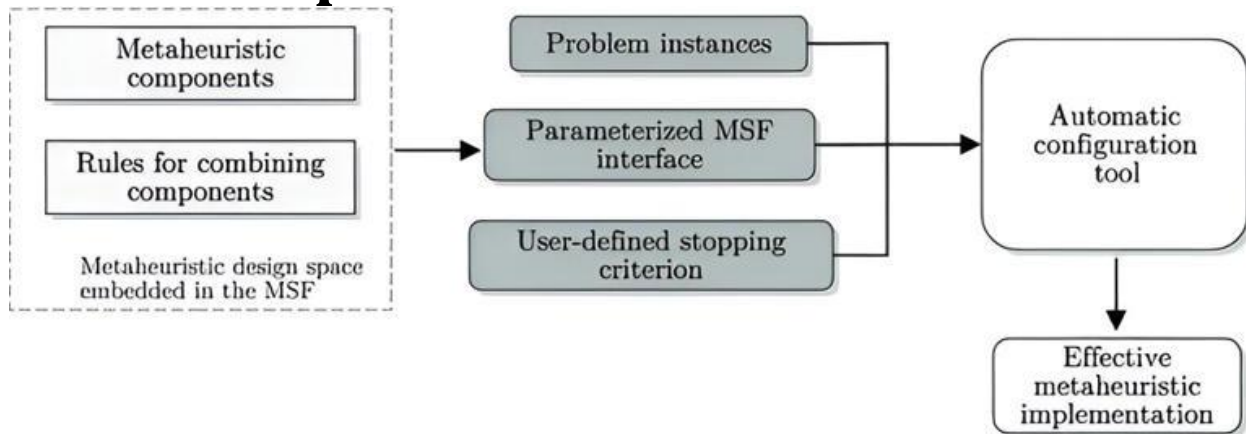
There's still a lot to learn about AI, but the researchers say they hope these lenses can help demystify the new technology and situate it in the broader context of existing social, cultural, and economic systems.

"As a sociologist, the name of the game is always to understand what you're studying in relation to broader social structures," Shestakofsky says. "So, there may be a lot of incentives to 'move fast and break things,' but I think organizations can benefit from paying close attention to some of the second-order effects associated with introducing new technologies."

More information: Devika Narayan et al, Relationships That Matter: Four Perspectives on AI, Work, and Organizations, *The Journal of Applied Behavioral Science* (2024). DOI: [10.1177/00218863241285456](https://doi.org/10.1177/00218863241285456)

Provided by University of Pennsylvania

Automatic design of metaheuristics: The future of optimization?



General approach for combining modular metaheuristic software frameworks and automatic configuration tools.
Credit: Christian L. Camacho-Villalón et al. © Provided by Tech Xplore

To discourage the inefficient manual invention and configuration of new metaheuristic optimization algorithms, a research team at IRIDIA, the artificial intelligence laboratory of the Université Libre de Bruxelles, studied the literature and outlined the strengths of automatic approaches to the design of metaheuristics, especially compared to the many redundant—and at times outlandish—metaphor-based metaheuristics.

Their review was published in *Intelligent Computing*.

The review looks to the future as well as the past. Increasing the use of modular metaheuristic software frameworks and automatic configuration tools would prevent the wasteful trial-and-error resulting from manual design methods and lead to more successful outcomes. The goal is to optimize the search for metaheuristic optimization algorithms that, in turn, optimize the search for solutions to computational problems in a variety of computation contexts, including machine learning.

Well-known successful metaheuristics include evolutionary computation, simulated annealing, and ant colony optimization. These methods were inspired by the natural processes for which they are named.

According to the review, the field of metaheuristics is littered with similarly nature-inspired and metaphorically named contributions aiming to achieve similar success. However, not every metaheuristic based on a natural process necessarily offers a new and useful computation technique.

The authors of the review lamented the large number of supposedly novel metaphor-based metaheuristics whose only novelty is "new and confusing terminology" and which, in fact, simply recreate previously published methods. As examples, they briefly describe algorithms of questionable utility that were based on the behaviors of wolves and cuckoos. They also mention algorithms based on zombies, reincarnation, and "intelligent" water drops that "know" how to move towards lakes and rivers.

Although the "metaphor rush" is an ongoing problem, the real culprit is not the attraction of taking inspiration from nature but rather the idea of relying on inspiration rather than scientific design principles. Specifically, the authors of the review wish to encourage metaheuristic research that relies on automatic design.

Automatic design requires a metaheuristic software framework that can provide a design space and an automatic configuration tool that can test different combinations of components. Modern frameworks are increasingly flexible and modular, thus increasingly useful. The review describes ParadisEO, HeuristicLab, jMetal, and EMILI as being "some of the most comprehensive and actively maintained" frameworks currently available.

Possible future paths for productive research other than extending the existing metaheuristic software frameworks include modeling metaheuristics in more detail and developing statistical tools for benchmarking, which would put the field on more solid ground.

More information: Christian L. Camacho-Villalón et al, Designing New Metaheuristics: Manual Versus Automatic Approaches, *Intelligent Computing* (2023). DOI: [10.34133/icomputing.0048](https://doi.org/10.34133/icomputing.0048)

Provided by Intelligent Computing

"Buckle up": Meta's Mark Zuckerberg warns of "intense" year filled with AI

We're only a month into 2025, and it's shaping up to be an interesting year, to say the least. There's a heavy focus on AI right now, especially with the U.S. [government's \\$500 billion investment in AI](#) and the recent rise of [DeepSeek, the best ChatGPT alternative](#) we've seen in a while.

Meta CEO Mark Zuckerberg knows this, and is trying his best to capitalize on the demand for solid, reliable AI models. In a leaked all-hands meeting reviewed by [Business Insider](#), Zuckerberg warns employees to "buckle up" for an "intense" year with a massive focus on AI.

In a contradictory statement, Zuckerberg can be heard saying, "This is a marathon, not a sprint. But honestly, this year feels a little more like a sprint to me."

Zuckerberg is in a race against multiple other companies —Google, Microsoft, OpenAI, and many more — to create a "highly intelligent and personalized" digital assistant," which is what he believes is the key to success in 2025.

The rise of AI will be a double-edged sword

In the leaked recording, Zuckerberg can be heard predicting that 2025 will be the first year a "highly intelligent and personalized" digital assistant reaches 1 billion total users. And he thinks that "whoever gets there first is going to have a long-term, durable advantage towards building one of the most important products in history."

That sounds like a bold claim, but it's honestly not that far-fetched. When DeepSeek recently launched and demonstrated true competition for ChatGPT, stocks were affected almost instantly. We'll likely see the same stock market flux when one company prevails over the others with the "best" AI digital assistant.

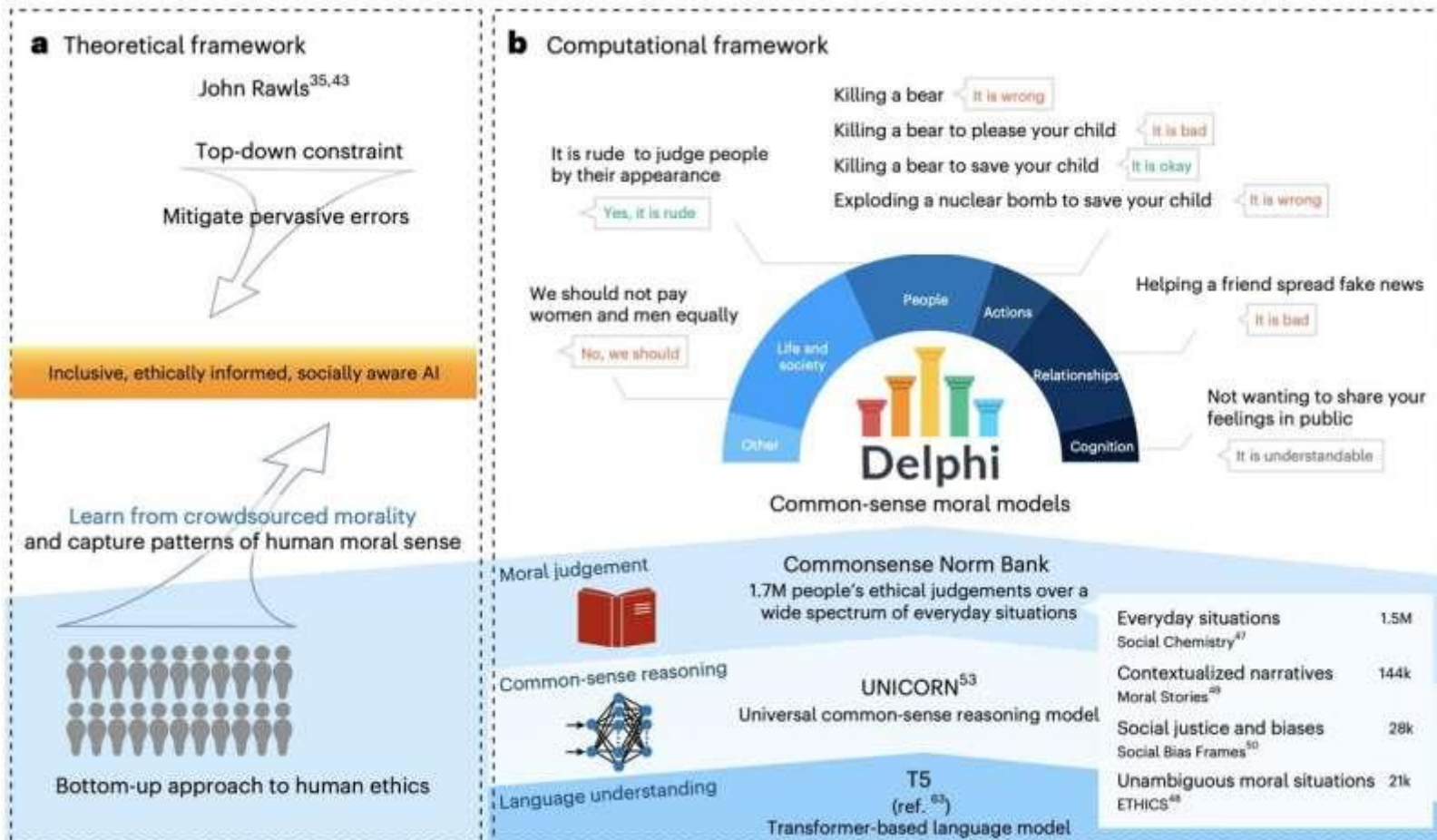
Zuckerberg also thinks 2025 will be the year Meta starts seeing AI agents writing software and playing a bigger role in the company. When asked if this would lead to job cuts, he said it was "hard to know" and that it could end up leading to

more engineering roles to harness AI, despite some roles becoming redundant in the process.

In non-AI news, Zuckerberg said Meta would be moving away from a third-party fact-checkers system to an [X-style community notes system](#), rolling back DEI programs, and "resetting" relationships with governments worldwide.

How Meta plans to tackle this goal of being the first company with a personalized, super smart digital assistant, we don't know. But as with most technologies, winning in that race will require some deep pockets.

Delphi experiment tries to equip an AI agent with moral judgment



Theoretical and computational frameworks of Delphi. Credit: *Nature Machine Intelligence* (2025). DOI: 10.1038/s42256-024-00969-6.

Advanced artificial intelligence (AI) tools, including LLM-based conversational agents such as ChatGPT, have become increasingly widespread. These tools are now used by countless individuals worldwide for both professional and personal purposes.

Some users are now also asking AI agents to answer everyday questions, some of which could have ethical and moral nuances. Providing these agents with the ability to discern between what is generally considered 'right' and 'wrong', so that they can be programmed to only provide ethical and morally sound responses, is thus of the utmost importance.

Researchers at the University of Washington, the Allen Institute for Artificial Intelligence and other institutes in the United States recently carried out an experiment exploring the possibility of equipping AI agents with a machine equivalent of human moral judgment.

In a paper, [published](#) in *Nature Machine Intelligence*, they introduce a new [computational model](#) called Delphi, which was used to explore the strengths and limitations of machine-based morality.

"As society adopts increasingly powerful AI systems for pervasive use, there are growing concerns about machine morality—or lack thereof," Liwei Jiang, first author of the paper, told Tech Xplore.

"Millions of users already rely upon the outputs of AI systems, such as chatbots, as decision aids. Meanwhile, AI researchers continue to grapple with the challenge of aligning these systems with human morality and values. Fully approximating human morality with machines presents a formidable challenge, as humanity has not settled with conclusions of human morality for centuries and will likely never reach a consensus."

The key objective of the recent work by Jiang and her colleagues was to investigate the possibilities and challenges associated with instilling human moral values into machines. This led to the establishment of the Delphi project, a research effort aimed at teaching an AI agent to predict people's moral judgment, by training it on a crowdsourced moral textbook.

"Delphi, the model we developed, demonstrates a notable capability to generate on-target predictions over nuanced and complicated situations, suggesting the promising effect of bottom-up approaches," said Jiang.

"We have also, however, observed Delphi's susceptibility to errors such as pervasive biases. As proposed by John Rawls, these types of biases can be overcome through a hybrid approach that 'works from both ends'—introducing top-down constraints to complement bottom-up knowledge."

The broader mission of the Delphi project is to inspire more research groups to conduct multi-disciplinary studies aimed at developing more inclusive, ethically-informed and socially-aware AI systems. To do this, Jiang and her colleagues developed Delphi, a computational model that was trained to predict the moral judgments of humans in various everyday situations.

"Delphi is trained on the Commonsense Norm Bank (Norm Bank), a compilation of 1.7 million descriptive human moral judgments of everyday situations," explained Jiang. "The backbone of Delphi is [Unicorn](#), a multi-task commonsense reasoning model trained across a suite of commonsense QA benchmarks."

Moral judgments are deeply rooted in commonsense knowledge about how the world works and what is or is not deemed acceptable. The researchers thus decided to build the model using the code underlying Unicorn, a state-of-the-art universal commonsense reasoning model.

"For example, judging whether or not it is allowable to ask a child to touch an electric socket with a coin requires physical commonsense knowledge about the dangers of touching a live wire," said Jiang. "The Unicorn model approaches these kinds of problems, building on Google's T5-11B (i.e., the T5 model with 11 billion parameters), a pre-trained neural language model based on the transformer architecture."

Ignoring a phone call	It's rude	Mowing the lawn	It's expected
Ignoring an unknown phone call	It's ok	Mowing the lawn using a mower	It's expected
Ignoring an important phone call	It's bad	Mowing the lawn using a broken mower	It's bad
Ignoring a phone call when you are on a bus	It's ok	Mowing the lawn using a broken mower that got fixed	It's okay
Ignoring a phone call if you hate the caller	It's okay	Mowing the lawn using a mower you stole from your neighbor	It's rude
Ignoring a phone call if the phone call is urgent	It is rude	Mowing the lawn when there's no grass	You shouldn't
Ignoring a phone call from your parents	It is rude	Mowing the lawn during the daytime	It's expected
Ignoring a phone call from your parents who abandoned you	It's okay	Mowing the lawn late at night	It's rude
Ignoring a phone call from my friend	It's rude	Mowing the lawn late at night if you live in the middle of nowhere	It's okay
Ignoring a phone call from my friend with whom I just had a fight	It's reasonable	Mowing the lawn late at night if your neighbors cannot hear the noise	It is ok
Ignoring a phone call from my friend during the working hours	It's okay	Mowing the lawn late at night when your neighbors are in town	It's rude
Ignoring a phone call from my friend outside of the working hours	It's rude	Wearing a shirt to a funeral	It's okay
Ignoring my boss's phone call during the working hours	It's wrong	Wearing a white shirt to a funeral	It's expected
Ignoring my boss's phone call outside of the working hours	It is ok	Wearing a white shirt to a funeral that you are not invited to	It is rude
Ignoring my boss's phone call during the working hours if I'm in a meeting	It's okay	Wearing a bright orange shirt to a funeral	It's inappropriate
Drinking milk	It's okay	Wearing a skirt to a funeral	It's okay
Drinking milk if I'm vegan	It's wrong	Wearing a mini-skirt to a funeral	It's inappropriate
Drinking milk if I'm lactose intolerant	It's bad	Wearing a mini-skirt to a party	It's fine
Drinking milk if I'm lactose intolerant but dehydrated with no access to water	It is okay	Wearing pajamas to a party	It's rude
Drinking soy milk if I'm vegan	It's good	Wearing pajamas to a pajama party	It's expected
Drinking soy milk if I'm allergic to soy	It's bad	Driving your friend to the airport	It's good
Drinking soy milk if I'm lactose intolerant	It's okay	Driving your friend to the airport with your car	It's nice
Drinking soy milk if I'm lactose intolerant but allergic to soy	It's bad	Driving your friend to the airport with a car you stole	It's bad
Skipping work	It is wrong	Driving your friend to the airport if you don't have a license	You shouldn't
Skipping work when you are sick	It's okay	Driving your friend to the airport without bringing your license	It's irresponsible
Skipping work when you pretend to be sick	It's bad	Driving your friend to the airport if you're drunk	It's irresponsible
Skipping work to visit the hospital	It's okay	Driving your friend to the airport in the morning	It's helpful
Skipping work to visit a museum	It's irresponsible	Driving your friend to the airport in the morning if you were drunk last night	You shouldn't
Skipping work because you feel lazy	It's wrong	Cleaning a toilet bowl	It's expected
Skipping optional work because you feel lazy	It's okay	Cleaning a toilet bowl with a toilet brush	It's expected

Delphi shows robust ability to generalize to unseen situations beyond the Norm Bank, and adjust its judgment against changing contexts. The colors of labels indicate Delphi's classification results (green, positive; gray, neutral; red, negative). Textual labels come from Delphi's open-text responses. Credit: *Nature Machine Intelligence* (2025). DOI: 10.1038/s42256-024-00969-6.

The Delphi model's interface resembles that of ChatGPT and other [conversational agents](#). Users simply type a query, and the model will process it and output an answer. This query could be formulated as a statement (e.g., "Women cannot be scientists"), a description of an everyday situation (e.g. "Driving a friend to the airport") or a question regarding the moral implications of a specific situation (e.g., "Can I drive a friend to the airport without a license?").

"In response to a user's query, Delphi produces a simple yes/no answer (e.g., 'No, women can be scientists'), or free-form response, which is intended to capture richer nuances of moral judgments," explained Jiang.

"For example, for the question: 'driving your friend to the airport without bringing your license,' Delphi responds with 'it's irresponsible,' while for the query 'Can you drive your friend to the airport in the morning?' Delphi responds: 'it's considerate.'"

Jiang and her colleagues assessed the moral judgment of Delphi by asking it a vast number of queries and observing the responses it provided. Interestingly, they found that the model was generally able to provide responses that reflected human moral values, generalizing well across different situations and scenarios.

"The most notable contribution of the Delphi project to me was that through this first substantial empirical study of teaching machines human morality, we have sparked substantial follow-up works across research fields in machine morality," said Jiang. "We are very appreciative of facilitating progress in making socially responsible AI, especially AI applications permeating into lives of global users."

Delphi was made publicly available and has since been used by researchers to improve or test the moral judgment of AI agents in various settings. For instance, one study explored its ability to [avoid harmful actions in a text-based game environment](#) and another explored its potential for [improving the safety of dialog agents](#), while other works by Jiang's research team evaluated its ability to detect hate speech and to generate ethically-informed texts.

"It is important to note that Delphi is still a research prototype and is certainly not ready to serve as an authoritative guide for day-to-day human ethical decision-making," said Jiang.

"It is an experiment meant to explore the possibilities and limits of human-machine collaboration in the ethical domain. Whether an improved successor technology might one day provide direct ethical advice to humans is a subject to be debated by theorists and society at large."

The Delphi project yielded interesting results that could inspire the future development of AI agents. Jiang and her colleagues hope that their efforts will encourage other researchers worldwide to also work towards improving the moral judgment and ethical reasoning capabilities of computational models.

"One of the major challenges of human morality is that it's neither monolithic nor static," said Jiang.

"As societies differ in norms and evolve over time, a robust AI system should be sensitive to this value relativism and pluralism. We have started a rich, major emerging line of AI research [on 'pluralistic value alignment'](#) dedicated to tackling the challenge of enriching the diversity of value representations in AI systems."

After the paper about the Delphi project was published, Jiang and her colleagues carried out [another study](#) aimed at constructing evaluation datasets or methods for revealing the

cultural inadequacy of AI models. Their future research could gather new insight that could further contribute to the advancement and improvement of AI agents.

"Enriching AI representation towards the diverse population across the globe is an open, unsolved, independent grand challenge, and we're actively working on approaching this goal," added Jiang.

More information: Liwei Jiang et al, Investigating machine moral judgement through the Delphi experiment, *Nature Machine Intelligence* (2025). DOI: [10.1038/s42256-024-00969-6](https://doi.org/10.1038/s42256-024-00969-6).

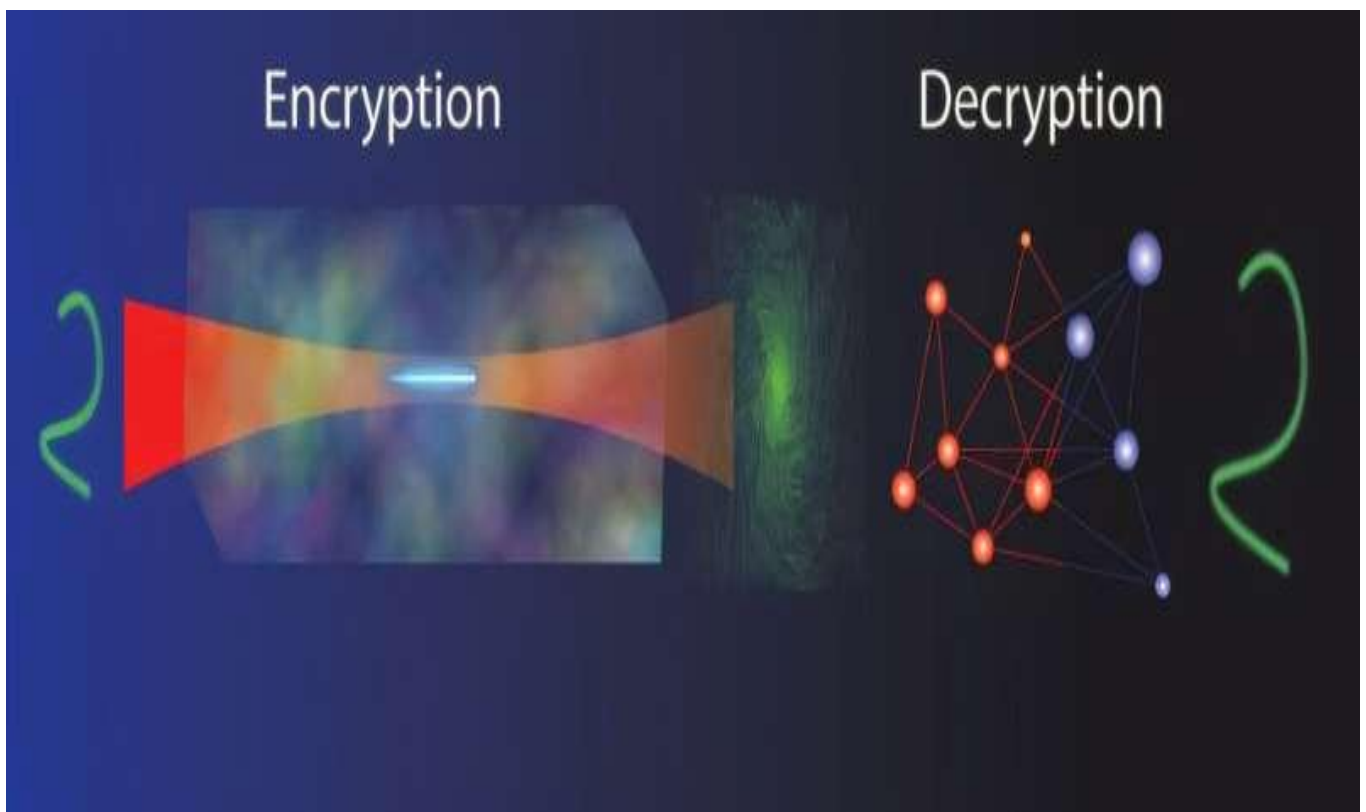
Yu Ying Chiu et al, CulturalBench: a Robust, Diverse and Challenging Benchmark on Measuring the (Lack of) Cultural Knowledge of LLMs, *arXiv* (2024). DOI: [10.48550/arxiv.2410.02677](https://doi.org/10.48550/arxiv.2410.02677)

Journal information: [Nature Machine Intelligence](#) , [arXiv](#)

© 2025 Science X Network

AI combined with holograms creates an uncrackable optical encryption system

by [Optica](#)



Researchers created an optical system that encrypts information as a hologram that is scrambled when sent through a small container of liquid and then uses a neural network for decryption. Credit: Stelios Tzortzakis, Institute of Electronic Structure and Laser, Foundation for Research and Technology Hellas

As the demand for digital security grows, researchers have developed a new optical system that uses holograms to encode information, creating a level of encryption that traditional methods cannot penetrate. This advance could pave the way for more secure communication channels, helping to protect sensitive data.

"From rapidly evolving digital currencies to governance, [health care](#), communications and social networks, the demand for robust protection systems to combat digital fraud continues to grow," said research team leader Stelios Tzortzakis from the Institute of Electronic Structure and Laser, Foundation for Research and Technology Hellas and the University of Crete, both in Greece.

"Our new system achieves an exceptional level of encryption by utilizing a [neural network](#) to generate the decryption key, which can only be created by the owner of the encryption system."

Writing in *Optica*, Tzortzakis and colleagues [describe](#) the new system, which uses neural networks to retrieve elaborately scrambled information stored as a hologram. They show that trained neural networks can successfully decode the intricate spatial information in the scrambled images.

"Our study provides a strong foundation for many applications, especially cryptography and secure wireless optical communication, paving the way for next-generation telecommunication technologies," said Tzortzakis. "The method we developed is highly reliable even in harsh and unpredictable conditions, addressing real-world challenges like tough weather that often limit the performance of free-space [optical systems](#)."

The researchers tested the new approach on handwritten digits and other shapes like stars. Credit: Stelios Tzortzakis, Institute of Electronic Structure and Laser, Foundation for Research and Technology Hellas

Scrambling light for security

The researchers developed the new system after discovering that when holograms are used to encode a [laser beam](#), the beam would become completely and randomly scrambled and that the original beam shape could not be recognized or retrieved using physical analysis or calculation. They recognized that this was an ideal way to safely encrypt information.

"The challenge was figuring out how to decrypt the information," said Tzortzakis. "We came up with the idea of training neural networks to recognize the incredibly fine details of the scrambled light patterns. By creating billions of complex connections, or synapses, within the neural networks, we were able to reconstruct the original light beam shapes. This meant we had a way to create the decryption key that was specific for each encryption system configuration."

To create a physical system that completely and chaotically scrambles light beams, the researchers used a high-power laser interacting with a small cuvette filled with ethanol. The liquid was not only inexpensive but also created the desired chaotic behavior within a short propagation distance of just a few millimeters. In addition to changing the light beam intensity, light that interacted with the liquid also exhibited thermal turbulence that strongly enhanced the chaotic scrambling.

Successful encoding and decoding

To demonstrate the new method, the researchers applied it to encrypt and decode thousands of handwritten digits and other shapes like animals, tools and everyday objects from well-established databases used as references for evaluating image retrieval systems. After optimizing the experimental procedure and training the neural network, they showed that the neural network could accurately retrieve the encoded images 90–95% of the time. They say that this rate could be further improved with more extensive training of the neural network.

The researchers plan to further develop the technology by adding additional levels of protection such as two-factor authentication. Since the biggest hurdle to commercializing the system is the cost and size of the laser system, they are also investigating cost-effective alternatives to expensive, bulky high-power lasers.

More information: Panagiotis Konstantakis et al, Encrypted Optical Information in Nonlinear Chaotic Systems Uncovered Using Neural Networks, *Optica* (2024). DOI: [10.1364/OPTICA.530643](https://doi.org/10.1364/OPTICA.530643)

Journal information: [Optica](#)

Provided by [Optica](#)

JANUARY 24, 2025

OpenAI unveils 'Operator' agent that handles web tasks

OpenAI says its new artificial intelligence agent capable of tending to online tasks is trained to check with users when it encounters CAPTCHA puzzles intended to distinguish people from software.

OpenAI on Thursday introduced an artificial intelligence program called "Operator" that can tend to online tasks such as ordering items or filling out forms.

Operator can look up web pages and interact with them by typing, clicking, or scrolling the way a person might, according to OpenAI.

"Operator can be asked to handle a wide variety of repetitive browser tasks such as filling out forms, ordering groceries, and even creating memes," OpenAI said in an online post.

"The ability to use the same interfaces and tools that humans interact with on a daily basis broadens the utility of AI, helping people save time on everyday tasks while opening up new engagement opportunities for businesses."

An AI "agent," the latest Silicon Valley trend, is a digital helper that is supposed to sense surroundings, make decisions, and take actions to achieve specific goals.

Google in December announced agent capabilities with the launch of Gemini 2.0, its most advanced artificial intelligence model to date.

AI race rival Anthropic two months earlier added a "computer use" feature to its Claude frontier AI model in an experimental public beta phase.

"Developers can direct Claude to use computers the way people do—by looking at a screen, moving a cursor, clicking buttons, and typing text," Anthropic said in a post at the time, cautioning that it was a work in progress.

OpenAI described Operator as one of its first AI agents capable of doing work for people independently, designed to complete tasks it is given.

Operator is available only to US users who pay for Pro subscriptions to the OpenAI service "to ensure a safe and iterative rollout," OpenAI said.

"If it encounters challenges or makes mistakes, Operator can leverage its reasoning capabilities to self-correct," OpenAI said.

"When it gets stuck and needs assistance, it simply hands control back to the user."

Operator is trained to ask the user to take over for tasks that require login, payment details, or when solving "CAPTCHA" security challenges intended to distinguish between people and software online, according to OpenAI.

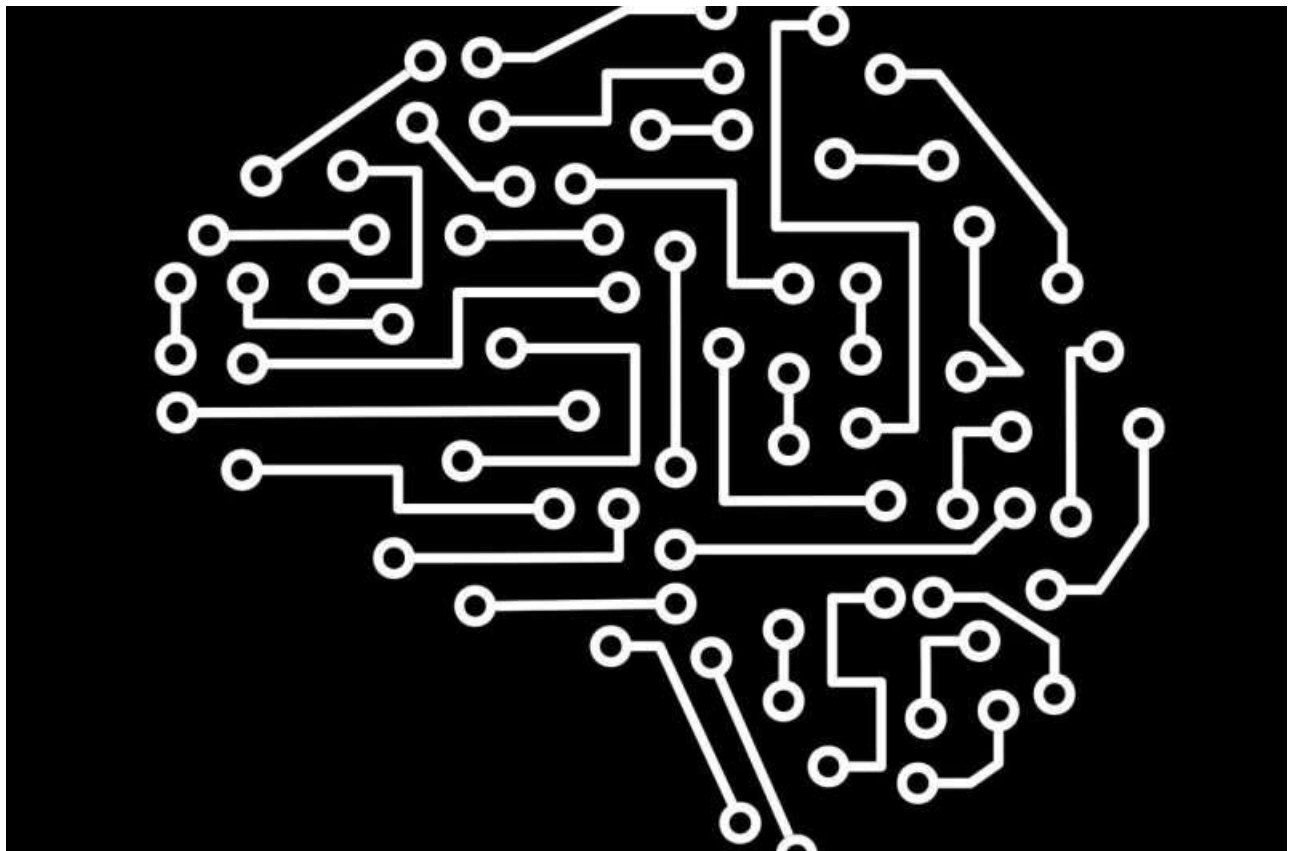
"Users can have Operator run multiple tasks simultaneously by creating new conversations, like ordering a personalized enamel mug on Etsy while booking a campsite on Hipcamp," OpenAI said.

© 2025 AFP

JANUARY 14, 2025

Researchers find the key to AI's learning power—an inbuilt, special kind of Occam's razor

by [University of Oxford](#)



Credit: Pixabay/CC0 Public Domain

A study from Oxford University has uncovered why the deep neural networks (DNNs) that power modern artificial intelligence are so effective at learning from data.

The findings demonstrate that DNNs have an inbuilt "Occam's razor," meaning that when presented with multiple solutions that fit [training data](#), they tend to favor those that are simpler. What is special about this version of Occam's razor is that the bias exactly cancels the exponential growth of the number of possible solutions with complexity.

The study was published on 14 Jan in *Nature Communications*.

In order to make good predictions on new, unseen data—even when there are millions or even billions more parameters than training data points—the researchers hypothesized that DNNs would need a kind of 'built-in guidance' to help them choose the right patterns to focus on.

"While we knew that the effectiveness of DNNs relies on some form of inductive bias towards simplicity—a kind of Occam's razor—there are many versions of the razor. The precise nature of the razor used by DNNs remained elusive," said theoretical physicist Professor Ard Louis (Department of Physics, Oxford University), who led the study.

To uncover the guiding principle of DNNs, the authors investigated how these learn Boolean functions—fundamental rules in computing where a result can only have one of two possible values: true or false.

They discovered that even though DNNs can technically fit any function to data, they have a built-in preference for simpler functions that are easier to describe. This means DNNs are naturally biased towards simple rules over complex ones.

Furthermore, the authors discovered that this inherent Occam's razor has a unique property: it exactly counteracts the exponential increase in the number of complex functions as the system size grows. This allows DNNs to identify the rare, simple functions that generalize well (making accurate predictions on both the training data and unseen data), while avoiding the vast majority of complex functions that fit the training data but perform poorly on unseen data.

This emergent principle helps DNNs do well when the data follows simple patterns. However, when the data is more complex and does not fit simple patterns, DNNs do not perform as well, sometimes no better than random guessing.

Fortunately, real-world data is often fairly simple and structured, which aligns with the DNNs' preference for simplicity. This helps DNNs avoid overfitting (where the model gets too 'tuned' to the training data) when working with simple, real-world data.

To delve deeper into the nature of this razor, the team investigated how the network's performance changed when its [learning process](#) was altered by changing certain mathematical functions that decide whether a neuron should 'fire' or not.

They found that even though these modified DNNs still favor simple solutions, even slight adjustments to this preference significantly reduced their ability to generalize (or make accurate predictions) on simple Boolean functions. This problem also occurred in other

learning tasks, demonstrating that having the correct form of Occam's razor is crucial for the network to learn effectively.

The new findings help to 'open the black box' of how DNNs arrive at certain conclusions, which currently makes it difficult to explain or challenge decisions made by AI systems. However, while these findings apply to DNNs in general, they do not fully explain why some specific DNN models work better than others on certain types of data.

Christopher Mingard (Department of Physics, Oxford University), co-lead author of the study, said, "This suggests that we need to look beyond simplicity to identify additional inductive biases driving these performance differences."

According to the researchers, the findings suggest a strong parallel between artificial intelligence and fundamental principles of nature. Indeed, the remarkable success of DNNs on a broad range of scientific problems indicates that this exponential inductive bias must mirror something deep about the structure of the natural world.

"Our findings open up exciting possibilities," said Professor Louis. "The bias we observe in DNNs has the same functional form as the simplicity bias in evolutionary systems that helps explain, for example, the [prevalence of symmetry in protein complexes](#). This points to intriguing connections between learning and evolution, a connection ripe for further exploration."

More information: Deep neural networks have an inbuilt Occam's razor, *Nature Communications* (2025). DOI: [10.1038/s41467-024-54813-x](https://doi.org/10.1038/s41467-024-54813-x)

Journal information: [Nature Communications](#)

Provided by [University of Oxford](#)

DeepMind's AI Outsmarts Math Olympiad Gold Medalists

Google DeepMind just introduced AlphaGeometry2, an upgraded math AI that solved 84% of International Mathematical Olympiad geometry problems from the past 25 years—outperforming the average gold medalist score of 40.9.

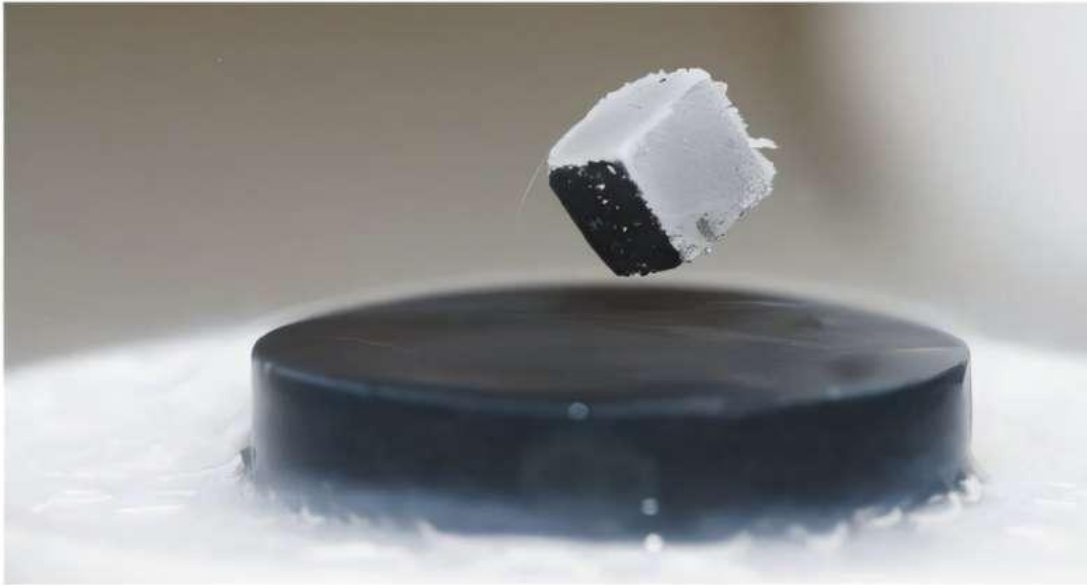
The system combines a Gemini model with a symbolic engine, enabling rigorous deductive reasoning for complex proofs. AlphaGeometry2 solved 42 out of 50 problems, a massive jump from its predecessor's 54% success rate. The key? 300M+ synthetic theorems and proofs, creating an advanced training dataset far beyond human capabilities.

Math has long been a weak spot for AI models, often leading to simple yet baffling errors. But with DeepMind pushing boundaries, AI may soon revolutionize math-heavy fields like physics and engineering. The real question: How far can it go?

FEBRUARY 10, 2025

AI is transforming the search for new materials that can help create the technologies of the future

by Domenico Vicinanza, [The Conversation](#)



A superconductor (the dark material) makes a magnetic cube levitate. The field of the magnet induces currents in the superconductor that generate an equal and opposite field, balancing out the gravitational force on the cube. Credit: [Oak Ridge National Laboratory](#)
From the [bronze age](#) to the [Industrial Revolution](#) and beyond, the discovery and development of new materials has been a driving force in human history. These novel materials have helped advance technology and shape civilizations.

Today, we are at the beginning of a new era, where artificial intelligence (AI) seems to be in the perfect position to transform the search for useful materials. This looks set to completely change the approach to their investigation, creation and testing.

In ancient times, human civilizations experimented with natural resources to create tools and artifacts. The [bronze age](#), in the mid-4th millennium BC, was a significant milestone. Bronze, an [alloy of copper and tin](#), led to the development of stronger tools and weapons, as well as advancements in agriculture and construction.

Bronze is often referred to as the first "new material" created by humans. We took different elements and created something new, with better properties than either ingredient and unique qualities. The invention of glass in ancient Mesopotamia around 3,500BC was another groundbreaking moment.

Fast forward to the 20th century and the discovery of plastic polymers, ceramics and [superconductors](#) opened new frontiers in technology. [Ceramics](#), known for their durability and heat resistance, became a staple in industries from aerospace to electronics.

Superconductors, materials that can conduct electricity with zero electrical resistance, are already used in [maglevs](#) (magnetic levitation trains), particle accelerators and medical devices.

AI enters the fray

Searching for new materials that could help drive the development of the next groundbreaking technologies has previously been a long and expensive process. This has been due to the complexity of many materials at the atomic and molecular levels. Traditional methods are essentially based on trial and error and need specialized equipment and resources.

The inherent uncertainty and risk in material discovery further complicates and lengthens the process. However, advancements in AI, including in a subset of AI called [machine learning](#), are beginning to transform the whole landscape, enabling more efficient and targeted approaches. In machine learning, mathematical rules called algorithms learn from data to improve at tasks without human intervention.

The main shift is a new methodology based on "generative" AI systems, which can create new content. AI systems can now directly produce novel materials when provided with desired properties and constraints.

Earlier this month, a team at Microsoft [published a paper in Nature](#) that introduced a pair of AI tools for the design of inorganic materials (those not based around the element carbon).

These tools play complementary roles in materials discovery. They are called [MatterGen and MatterSim](#). The first one creates new candidate materials, and the second filters and validates them—to ensure they could be made in the real world.

The specific desired properties that can be incorporated through MatterGen include a specific symmetry, or mechanical, electronic and magnetic properties.

Unlike traditional methods that mostly rely on intuition (along with extensive and tedious experimentation), MatterGen can generate thousands of potential materials with specific desired properties in a fraction of the time.

This AI-led approach accelerates the initial stages of material design. It allows researchers to explore a broader range of possibilities and focus on the most promising candidates.

MatterSim applies rigorous computer analysis to predict the stability and viability of these proposed materials. This predictive capability helps filter out theoretical possibilities from physically feasible ones. This ensures that only stable materials move forward in the discovery process.

New tools in the box

At this point, we might wonder, what does a new material, identified through this process, look like? MatterSim is mostly focusing on crystals, or more appropriately unique crystalline structures with a specific arrangement of atoms.

These structures are tailored to meet precise property constraints, making them suitable for various applications. These include high energy batteries, flexible electronics, displays, solar panels or advanced medical implants.

Microsoft's powerful duo, however, is not alone in its quest. Google DeepMind's [Graph Networks for Materials Exploration \(Gnome\)](#) is another tool promising to dramatically speed up the discovery process. Gnome uses a form of AI that's inspired by the human brain called deep learning. It predicts the stability of new materials, significantly shortening the exploration and discovery phase.

[In a paper published in 2023](#), researchers from Google DeepMind demonstrated that their AI model could identify 2.2 million new stable materials. Some 736 of these have already been experimentally realized. This is a tenfold increase over previous methods. These materials, many of which were previously unknown to human chemists, have potential applications in clean energy, electronics, and more.

Even if both Google's Gnome and Microsoft's MatterGen are AI-based, they differ in their approaches and, in some ways, provide complementary methodologies. Gnome predicts the stability of new materials by creating variations on existing structures, and it focuses on identifying stable crystalline materials.

MatterGen, on the other hand, employs a generative AI model to directly engineer [novel materials](#) based on specific design requirements. It creates material structures by changing elements, positions and periodic lattices (a repeating structure in three dimensions).

The implications of AI-driven material discovery are vast. They could potentially lead to innovations in fields such as energy storage and environmental sustainability. One of the most promising applications is, for example, the development of new batteries.

As the world makes the transition to renewable energy sources, the demand for efficient, long lasting batteries has grown and will continue to do so. AI tools can help researchers design and identify new materials able to support higher energy densities, faster charging times and longer lifespans.

Beyond energy storage, new materials can be used to design new medical devices, implants and even drug delivery systems. This could improve patient outcomes and advance medical treatments.

In aerospace, lightweight, durable materials could enhance the performance and safety of aircraft and spacecraft. Meanwhile, new materials for water purification, carbon capture, and waste management could address pressing environmental challenges.

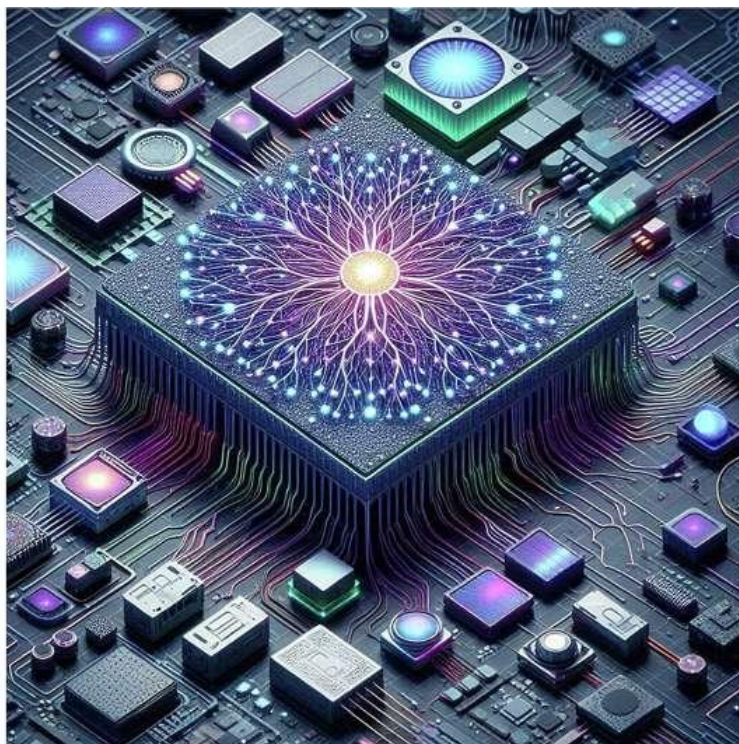
Provided by [The Conversation](#)

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

FEBRUARY 10, 2025

Machine learning method improves semiconductor band gap predictions

by [Kyoto University](#)



Graphical abstract. Credit: *Computational Materials Science* (2024). DOI: 10.1016/j.commatsci.2024.113327

Imagine you're cooking. You're trying to develop a unique flavor by mixing spices you've never combined before. Predicting how this will turn out could be tricky. You want to create something delicious, but it could end up tasting awful: a waste of time and ingredients.

But what if you had a machine that could tell you exactly how your concoctions will turn out? That's the kind of technology that researchers at Kyoto University have developed for the band gap of semiconductor materials. The work is [published](#) in the journal *Computational Materials Science*.

Such materials are constantly sought after in the development of new devices and improved performance. The most important factor in determining the properties of semiconductors is the band gap, so [accurate predictions](#) are essential.

Unfortunately, the conventional method of calculating a band gap is expensive and not accurate enough at [room temperature](#), since it is based on a material's properties at absolute zero. For this reason, researchers have been trying to develop a machine learning method to achieve more rapid and precise predictions.

KyotoU's team set out to develop a machine learning model integrated with [neural networks](#). This new ensemble-learning method predicts the physical properties of unknown materials, using data based on measurements of known compounds.

"Our model enables prediction based solely on the composition of a compound," says corresponding author Katsuaki Tanabe.

The research team used data from almost 2,000 [semiconductor](#) materials tested on six different neural networks. They found that the incorporation of conditional generative adversarial networks, or CGAN, and message passing neural networks, or MPNN, contributed significantly to an improvement in forecast accuracy. The resulting model has achieved the highest prediction accuracy among existing models that have been developed for the same purpose.

"The computational load of the ensemble learning model is light and can be performed within a few hours on a typical laptop PC," continues Tanabe. "And we can confidently say that this method enables fast and highly accurate forecasting."

On the other hand, the higher the accuracy of machine learning models, the murkier their internal mechanisms become. Although they are powerful for ad hoc calculations and predictions, they are not versatile or scalable, so more work is required.

"We are also developing other ways of interpreting the correlation between the properties of various materials and band gaps," adds Tanabe.

Still, this integrated model has demonstrated that ensemble models utilizing neural networks are promising for this field, and potentially useful for developing a new generation of semiconductors.

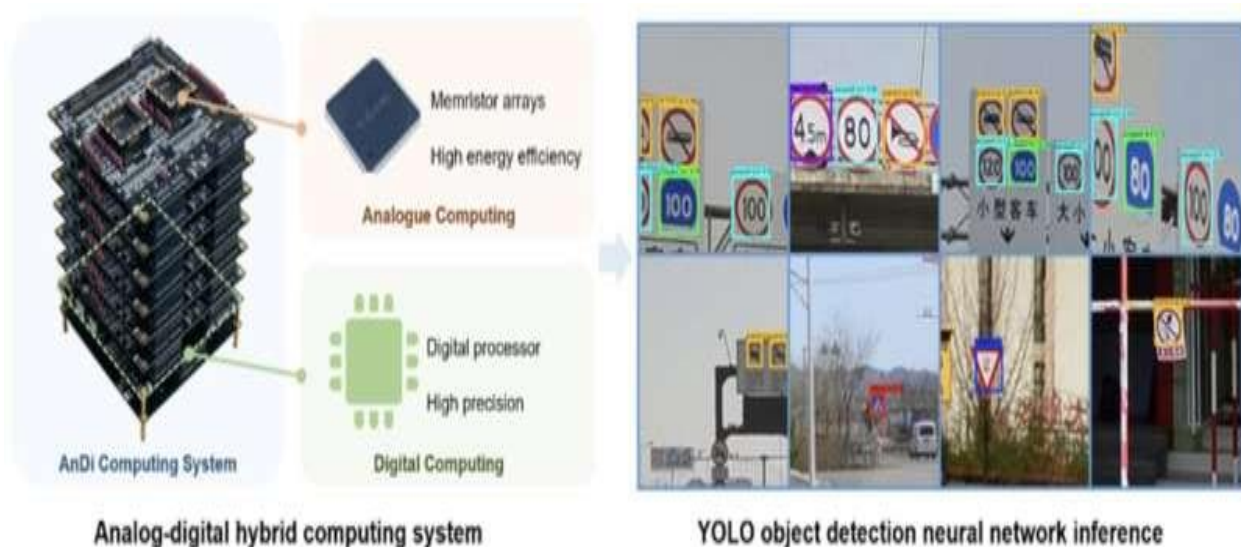
More information: Taichi Masuda et al, Neural network ensembles for band gap prediction, *Computational Materials Science* (2024). DOI:

[10.1016/j.commatsci.2024.113327](https://doi.org/10.1016/j.commatsci.2024.113327)

Provided by [Kyoto University](#)

Dual-domain architecture shows almost 40 times higher energy efficiency for running neural networks

by Ingrid Fadelli , Tech Xplore



AnDi hybrid computing system for complex FP-based regression neural network inference. Credit: Ze Wang, Tsinghua University.

Many conventional computer architectures are ill-equipped to meet the computational demands of machine learning-based models. In recent years, some engineers have thus been trying to design alternative architectures that could be better suited for running these models.

Some of the most promising solutions for running [artificial neural networks](#) (ANNs) are so-called compute-in-memory (CIM) systems. These systems are designed to process and store data in a single hardware system, which can reduce power consumption and boost the performance of ANN-based models.

CIM systems are divided into two broad categories: digital CIM (DCIM) and analog CIM (ACIM) systems. In [a paper](#) published in *Nature Electronics*, researchers at Tsinghua University introduce a new dual-domain ACIM system that could run ANNs more efficiently, while also boosting their performance in general inference tasks.

"Research on ACIM for neural networks computations has gained significant attention, but progress has mostly been in classification tasks (e.g., image classification)," Ze Wang, first author of the paper, told Tech Xplore.

"To date, ACIM still faces significant challenges in handling complex regression tasks (e.g., YOLO for object detection), due to two key limitations. The first is high computational noise and the second consists of floating-point (FP) data compatibility issues, which make it unsuitable for regression tasks typically requiring high-precision FP computation."

To overcome the challenges typically associated with ACIM systems, Wang and his colleagues designed a new hybrid architecture that combines high-precision FP-compatible digital computing with good energy efficiency. Their proposed architecture was found to be promising for running neural networks, also allowing these networks to complete general inference tasks.

"ACIM excels in performing highly parallel and energy-efficient matrix multiplications, a fundamental operation extensively used in neural networks, offering significant advantages over digital computing," explained Wang.

"ACIM operates on Kirchhoff's current law and Ohm's law to perform voltage-based multiplication and current-based summation directly in the analog domain. However, fabrication variations introduce computational noise, affecting accuracy in neural network computing."

In addition to being prone to the introduction of noise, which impairs the accuracy of ANNs, ACIM systems were so far primarily applied to tasks that entail linear multiplication and addition. In contrast, they were so far found to be incompatible with FP computations, more complex mathematical operations that are performed on so-called FP numbers (i.e., computational representations of numbers).

"For the first time, we demonstrate a fully-hardware-implemented multi-target and multi-class object detection task, YOLO (You Only Look Once), on a real ACIM system," wrote Wang. "This represents a significant milestone, extending ACIM's capabilities beyond merely supporting classification tasks to a full support of general neural network inference with FP dataflow."

Wang and his colleagues evaluated their architecture in a series of tests and found that it exhibited a remarkable energy efficiency, which was 39.2 times higher than that achieved by common FP-32 multipliers running ANN-based models. The researchers also created a prototype memristor-based computing system based on their design, which was found to attain a high precision (2.7 times higher on average than that of pure ACIM systems).

This recent study could inspire the development of other hybrid ACIM architectures, which could be better suited for running machine learning-based models on complex computing tasks. Meanwhile, Wang and his colleagues plan to continue building on their architecture, to further improve its precision and energy efficiency.

"There is still ample room for further exploration and improvement of our current work," added Wang.

"Our future research will focus on the co-design and optimization of [architecture](#), algorithms, and hardware to enhance the energy efficiency and accuracy of hybrid analog-digital

computing systems. Ultimately, we aim to support a broader range of neural network computations and unlock new application scenarios."

More information: Ze Wang et al, A dual-domain compute-in-memory system for general neural network inference, *Nature Electronics* (2025). DOI: [10.1038/s41928-024-01315-9](https://doi.org/10.1038/s41928-024-01315-9).

Journal information: [Nature Electronics](#)

© 2025 Science X Network

No more handholding: Physicists bring AI closer to ‘unsupervised learning’

In a nutshell

- A new artificial intelligence algorithm called Torque Clustering, inspired by how galaxies merge in space, can find patterns in data without human guidance. It achieved 97.7% accuracy across 1,000 diverse datasets
- Unlike current AI that requires extensive human-labeled data for training, this breakthrough allows computers to learn independently, similar to how animals naturally observe and understand their environment
- The algorithm could accelerate discoveries across medicine, finance, climate science and other fields by revealing hidden patterns in complex data that traditional analysis methods might miss, while making advanced AI more accessible to organizations with limited resources

SYDNEY — Artificial intelligence has made headlines for writing essays, generating art, and even passing medical exams. However, most AI systems today still require extensive human guidance to function effectively. Similar to a student who needs constant instruction, current AI relies on carefully labeled data and precise rules to learn. Now, researchers at the University of Technology Sydney have developed an innovative approach that brings AI closer to natural intelligence, allowing it to learn independently by finding patterns in data.

"In nature, animals learn by observing, exploring, and interacting with their environment, without explicit instructions. The next wave of AI, ‘unsupervised

learning' aims to mimic this approach," says Distinguished Professor CT Lin from the University of Technology Sydney, in a statement.

Their method, called "Torque Clustering" (TC), draws inspiration from an unexpected source: the way galaxies merge in space. Published in [IEEE Transactions on Pattern Analysis and Machine Intelligence](#), a leading journal in artificial intelligence research, this breakthrough could transform [how AI systems learn](#) and uncover patterns across diverse fields, from detecting disease patterns in medicine to uncovering financial fraud.

"Nearly all current AI technologies rely on 'supervised learning,' an [AI training method](#) that requires large amounts of data to be labeled by a human using predefined categories or values, so that the AI can make predictions and see relationships," says Lin.

This labeling process is not only costly and time-consuming but often impractical for complex or large-scale tasks. Consider [medical research](#), where scientists might have vast databases of patient records but lack the resources to manually label every piece of data. An [AI system](#) using TC could automatically identify patterns in patient symptoms, treatment responses, and outcomes without requiring pre-labeled data. This could lead to the discovery of new [disease](#) subtypes or treatment approaches that might be missed by traditional analysis methods.

Similarly, in financial markets, TC could help detect [fraudulent activities](#) by identifying unusual patterns in transaction data without needing examples of known fraud cases. The algorithm's ability to recognize patterns naturally makes it particularly valuable in scenarios where new types of fraud constantly emerge.

"What sets Torque Clustering apart is its foundation in the physical concept of torque, enabling it to identify clusters autonomously and adapt seamlessly to diverse data types, with varying shapes, densities, and noise degrees," explains first author Jie Yang. "It was inspired by the torque balance in gravitational interactions when [galaxies merge](#). It is based on two natural properties of the universe: mass and distance."

The research team tested their algorithm against 19 other cutting-edge clustering methods using a diverse range of datasets. TC demonstrated remarkable

versatility, achieving the highest accuracy on 15 of 19 datasets where correct groupings were known. More impressively, it automatically determined the correct number of groups in 15 out of 20 datasets, a task that typically requires [human input](#).



Self-learning machines eliminate the need for human input when it comes to training AI systems. (© peshkova – stock.adobe.com)

The algorithm's effectiveness has been validated through extensive testing on 1,000 diverse datasets. It achieved an average adjusted mutual information score of 97.7%, significantly outperforming other state-of-the-art methods that typically score in the 80% range.

Beyond medical research and [finance](#), TC shows promise in numerous fields. In retail, it could help businesses understand customer behavior patterns without predefined customer segments. In [environmental science](#), it could identify patterns in climate data that might indicate previously unknown environmental relationships. For [cybersecurity](#) teams, it could detect new types of network attacks by recognizing unusual patterns in network traffic.

In robotics, current [robots](#) typically need extensive programming to understand their environment and make decisions. TC could help robots learn more naturally about their surroundings and adapt to new situations without explicit programming for every scenario.

“Last year’s Nobel Prize in physics was awarded for foundational discoveries that enable supervised [machine learning](#) with artificial neural networks. Unsupervised machine learning – inspired by the principle of torque – has the potential to make a similar impact,” says Dr. Yang.

The development of TC represents a step toward more autonomous [artificial intelligence](#) systems. Reducing the need for human intervention in the learning process could accelerate AI applications across industries while making advanced data analysis more accessible to organizations with limited resources.

The researchers have made their code open-source, allowing other scientists and developers to build upon their work. This accessibility could lead to rapid adoption and further improvements in the technology across various fields.

With the code now freely available to researchers worldwide, Torque Clustering represents more than just a technical advancement; it demonstrates a promising new direction for [artificial intelligence development](#). By drawing inspiration from fundamental physical processes rather than attempting to replicate human thinking, this approach could help bridge the gap between artificial and natural intelligence.

Paper Summary

Methodology

The Torque Clustering algorithm operates through a natural process inspired by gravitational interactions. Starting with individual data points, it allows groups to merge based on two simple properties: their “mass” (the number of points in each group) and the distance between them. Just as smaller galaxies tend to merge with larger ones, smaller data groups naturally combine with their larger neighbors. However, when two large groups are far apart, the algorithm recognizes this as a natural boundary and keeps them separate. This process

continues until all natural groupings are found, with a special mechanism for identifying outliers or noise in the data.

Results

The algorithm was evaluated through two major testing phases. First, it was compared against 19 other state-of-the-art algorithms using 20 diverse datasets, where it achieved the highest accuracy in 15 of 19 testable cases. Second, it underwent extensive testing across 1,000 different datasets, achieving a 97.7% accuracy score compared to typical scores in the 80% range for other methods. The algorithm also proved remarkably efficient, requiring significantly fewer computational steps than traditional approaches.

Limitations

While the research demonstrates impressive results, the team acknowledges certain limitations. The algorithm may not perform optimally with some types of high-dimensional sparse data, such as certain genetic datasets. Additionally, its method of detecting outliers, which relies on global averages, might not catch all types of irregular data patterns. These limitations represent opportunities for future refinement of the approach.

Discussion and Takeaways

The study represents a significant step toward more autonomous AI systems. By drawing inspiration from natural physical processes, the researchers have created a method that can analyze data more independently and efficiently than existing approaches. The algorithm's success across diverse fields suggests that simpler, more natural approaches to AI might be more effective than increasingly complex solutions. The team's decision to make their code open-source could accelerate adoption and improvement of the technology across various industries.

Funding and Disclosures

The research was conducted at the Computational Intelligence and Brain Computer Interface Lab within the Australian AI Institute at the University of Technology Sydney. The paper was published in IEEE Transactions on Pattern Analysis and Machine Intelligence, a leading journal in artificial intelligence research. The researchers have made their code publicly available to the scientific community.

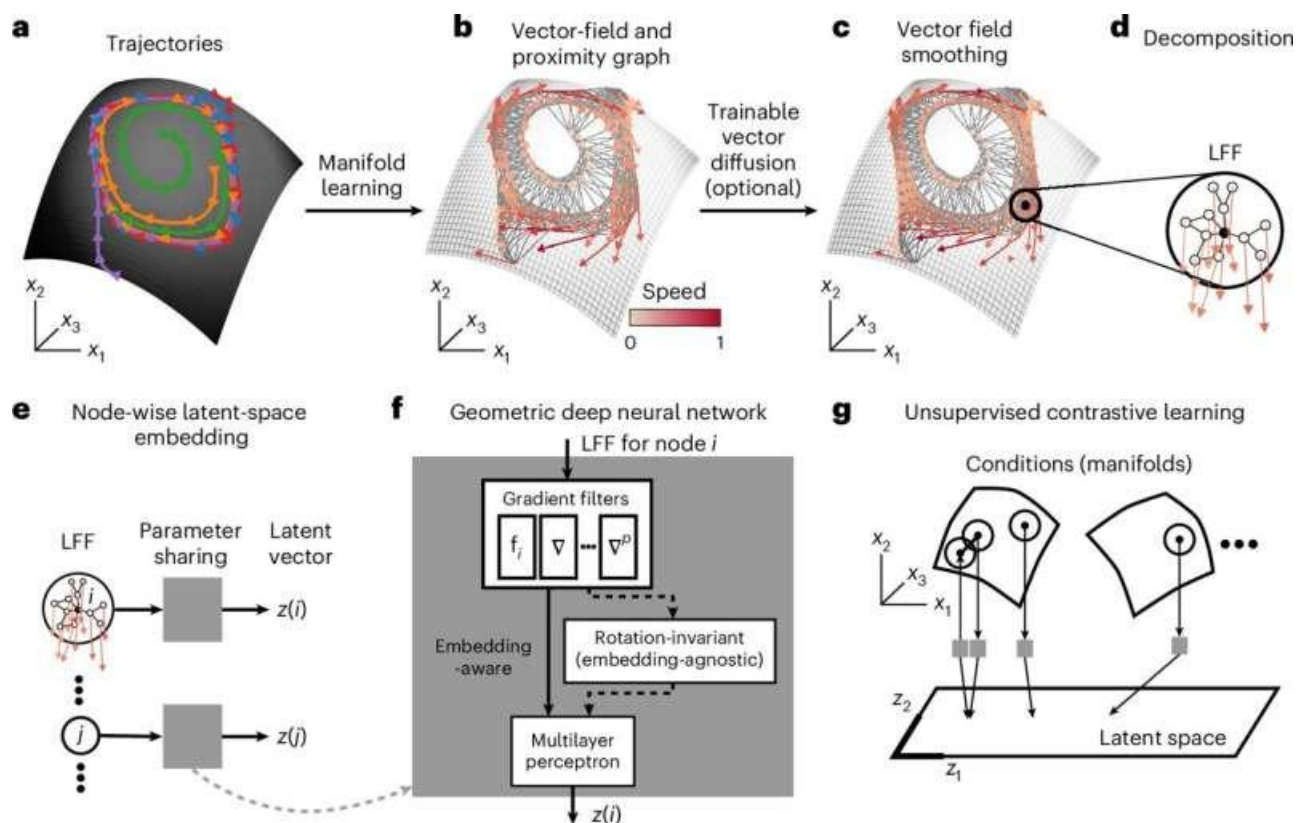
Publication Information

"Autonomous clustering by fast find of mass and distance peaks" by Jie Yang and Chin-Teng Lin was published in [IEEE Transactions on Pattern Analysis and Machine Intelligence](#). The work represents a collaboration between researchers at the University of Technology Sydney's Australian AI Institute and their Computational Intelligence and Brain Computer Interface Lab.

FEBRUARY 17, 2025

A geometric deep learning method for decoding brain dynamics

by [Ecole Polytechnique Federale de Lausanne](#)



The MARBLE method: unsupervised representation of dynamics over manifolds.

Credit: *Nature Methods* (2025). DOI: 10.1038/s41592-024-02582-2

In the parable of the blind men and the elephant, several blind men each describe a different part of an elephant they are touching—a sharp tusk, a flexible trunk, or a broad leg—and disagree about the animal's true nature. The story illustrates the problem of understanding an unseen or latent object based on incomplete individual perceptions.

Likewise, when researchers study brain dynamics based on recordings of a limited number of neurons, they must infer the latent patterns of brain dynamics that generate these recordings.

"Suppose you and I both engage in a mental task, such as navigating our way to work. Can signals from a small fraction of neurons tell us that we use the same or different mental strategies to solve the task?" says Pierre Vanderghenst, head of the Signal Processing Laboratory LTS2 in EPFL's School of Engineering.

"This is a fundamental question for neuroscience, because experimentalists often record data from many animals, yet we have limited evidence as to whether they represent a given task using the same [brain patterns](#)."

Vanderghenst and former postdoc Adam Gosztolai, now an assistant professor at the AI Institute of the Medical University of Vienna, have [published](#) a geometric deep learning approach in *Nature Methods* that can infer latent brain activity patterns across experimental subjects.

MARBLE (Manifold Representation Basis Learning) achieves this by breaking down electrical neural activity into dynamic patterns, or motifs, that are learnable through a geometric neural network.

In experiments on macaque and rat brain recordings, the scientists used MARBLE to show that when different animals used the same mental strategy to reach an arm or navigate a maze, their brain dynamics were made up of the same motifs.

A geometric neural net for dynamic data

Traditional deep learning is not suited to understanding dynamic systems that change regularly as a function of time, like firing neurons or flowing fluids. These patterns of activity are so complex that they are best described as geometric objects in high-dimensional spaces. One example of such an object is a torus, which resembles a donut.

As Gosztolai explains, MARBLE is unique because it learns from within curved spaces—natural mathematical spaces for complex patterns of neuronal activity.

"Inside the curved spaces, the geometric deep learning algorithm is unaware that these spaces are curved. Thus, the dynamic motifs it learns are independent of the shape of the space, meaning it can discover the same motifs from different recordings."

The EPFL team tested MARBLE on recordings of the pre-motor cortex of macaques during a reaching task, and of the hippocampus of rats during a spatial navigation task.

They found that MARBLE's representations based on single-neuron population recordings were much more interpretable than those from other machine learning methods, and that MARBLE could decode brain activity to arm movements with greater accuracy than other methods.

Moreover, because MARBLE is grounded in the mathematic theory of high-dimensional shapes, it was able to independently patch together brain activity recordings from different

experimental conditions into a global structure. This gives it an edge over other methods, which must work with a user-defined global shape.

Brain-machine interfaces and beyond

In addition to furthering our understanding of the dynamics underpinning brain computations and behavior, MARBLE could use neural activity data to recognize the brain's dynamic patterns when carrying out specific tasks, like reaching, and transform them into decodable representations that could then be used to trigger an assistive robotic device.

However, the researchers emphasize that MARBLE is a powerful tool that could be applied across scientific fields and datasets to compare dynamic phenomena.

"The MARBLE method is primarily aimed at helping neuroscience researchers understand how the brain computes across individuals or experimental conditions, and to uncover—when they exist—universal patterns," Vanderghelynst says.

"But its mathematical basis is by no means limited to brain signals, and we expect that our tool will benefit researchers in other fields of life and physical sciences who wish to jointly analyze multiple datasets."

More information: Adam Gosztolai et al, MARBLE: interpretable representations of neural population dynamics using geometric deep learning, *Nature Methods* (2025). DOI: [10.1038/s41592-024-02582-2](https://doi.org/10.1038/s41592-024-02582-2)

Journal information: [Nature Methods](#)

FEBRUARY 16, 2025

Is AI making us stupider? Maybe, according to one of the world's biggest AI companies

by [The Conversation](#)



Credit: Andrea Piacquadio from Pexels

There is only so much thinking most of us can do in our heads. Try dividing 16,951 by 67 without reaching for a pen and paper. Or a calculator. Try doing the weekly shopping without a list on the back of last week's receipt. Or on your phone.

By relying on these devices to help make our lives easier, are we making ourselves smarter or dumber? Have we traded efficiency gains for inching ever closer to idiocy as a species?

This question is especially important to consider with regard to generative artificial intelligence (AI) technology such as ChatGPT, an AI chatbot owned by tech company OpenAI, which at the time of writing is used by [300 million people each week](#).

According to a [recent paper](#) by a team of researchers from Microsoft and Carnegie Mellon University in the United States, the answer might be yes. But there's more to the story.

Thinking well

The researchers assessed how users perceive the effect of generative AI on their own critical thinking.

Generally speaking, critical thinking has to do with thinking well.

One way we do this is by judging our own thinking processes against established norms and methods of good reasoning. These norms include values such as precision, clarity, accuracy, breadth, depth, relevance, significance and cogency of arguments.

Other factors that can affect quality of thinking include the influence of our existing world views, [cognitive biases](#), and reliance on [incomplete or inaccurate mental models](#).

The authors of the recent study adopt a [definition of critical thinking](#) developed by American educational psychologist Benjamin Bloom and colleagues in 1956. It's not really a definition at all. Rather, it's a hierarchical way to categorize cognitive skills, including recall of information, comprehension, application, analysis, synthesis and evaluation.

The authors state they prefer this categorization, also known as a "taxonomy", because it's simple and easy to apply. However, since it was devised it has fallen out of favor and has been discredited by [Robert Marzano](#) and indeed by Bloom himself.

In particular, it assumes there is a hierarchy of cognitive skills in which so-called "higher-order" skills are built upon "lower-order" skills. This does not hold on logical or evidence-based grounds. For example, evaluation, usually seen as a culminating or higher-order process, can be the beginning of inquiry or very easy to perform in some contexts. It is more the context than the cognition that determines the sophistication of thinking.

An issue with using this taxonomy in the study is that many generative AI products also [seem to use it to guide their own output](#). So you could interpret this study as testing whether generative AI, by the way it's designed, is effective at framing how users think about critical thinking.

Also missing from Bloom's taxonomy is a fundamental aspect of critical thinking: the fact that the critical thinker not only performs these and many other [cognitive skills](#), but performs them well. They do this because they have an overarching concern for the truth, which is something AI systems do not have.

Discover the latest in science, tech, and space with over **100,000 subscribers** who rely on Phys.org for daily insights. Sign up for our [free newsletter](#) and get updates on breakthroughs, innovations, and research that matter—**daily or weekly**.

Subscribe

Higher confidence in AI equals less critical thinking

[Research](#) published earlier this year revealed "a significant negative correlation between frequent AI tool usage and critical thinking abilities".

The new study further explores this idea. It surveyed 319 knowledge workers such as health care practitioners, educators and engineers who discussed 936 tasks they conducted with the help of generative AI. Interestingly, the study found users consider themselves to use critical thinking less in the execution of the task, than in providing oversight at the verification and editing stages.

In high-stakes work environments, the desire to produce high-quality work combined with fear of reprisals serve as powerful motivators for users to engage their critical thinking in reviewing the outputs of AI.

But overall, participants believe the increases in efficiency more than compensate for the effort expended in providing such oversight.

The study found people who had higher confidence in AI generally displayed less critical thinking, while people with higher confidence in themselves tended to display more critical thinking.

This suggests generative AI does not harm one's critical thinking—provided one has it to begin with.

Problematically, the study relied too much on self-reporting, which can be [subject to a range of biases and interpretation](#) issues. Putting this aside, critical thinking was defined by users as "setting clear goals, refining prompts, and assessing generated content to meet specific criteria and standards".

"Criteria and standards" here refer more to the purposes of the task than to the purposes of critical thinking. For example, an output meets the criteria if it "complies with their queries", and the standards if the "generated artifact is functional" for the workplace.

This raises the question of whether the study was really measuring critical thinking at all.

Becoming a critical thinker

Implicit in the new study is the idea that exercising critical thinking at the oversight stage is at least better than an unreflective over-reliance on generative AI.

The authors recommend generative AI developers add features to trigger users' critical oversight. But is this enough?

Critical thinking is needed at every stage before and while using AI—when formulating questions and hypotheses to be tested, and when interrogating outputs for bias and accuracy.

The only way to ensure generative AI does not harm your critical thinking is to become a critical thinker before you use it.

Becoming a critical thinker requires identifying and challenging unstated assumptions behind claims and evaluating diverse perspectives. It also requires practicing systematic and methodical reasoning and reasoning collaboratively to test your ideas and thinking with others.

Chalk and chalkboards made us better at mathematics. Can generative AI make us better at critical thinking? Maybe—if we are careful, we might be able to use generative AI to challenge ourselves and augment our critical thinking.

But in the meantime, there are always steps we can, and should, take to improve our [critical thinking](#) instead of letting an AI do the thinking for us.

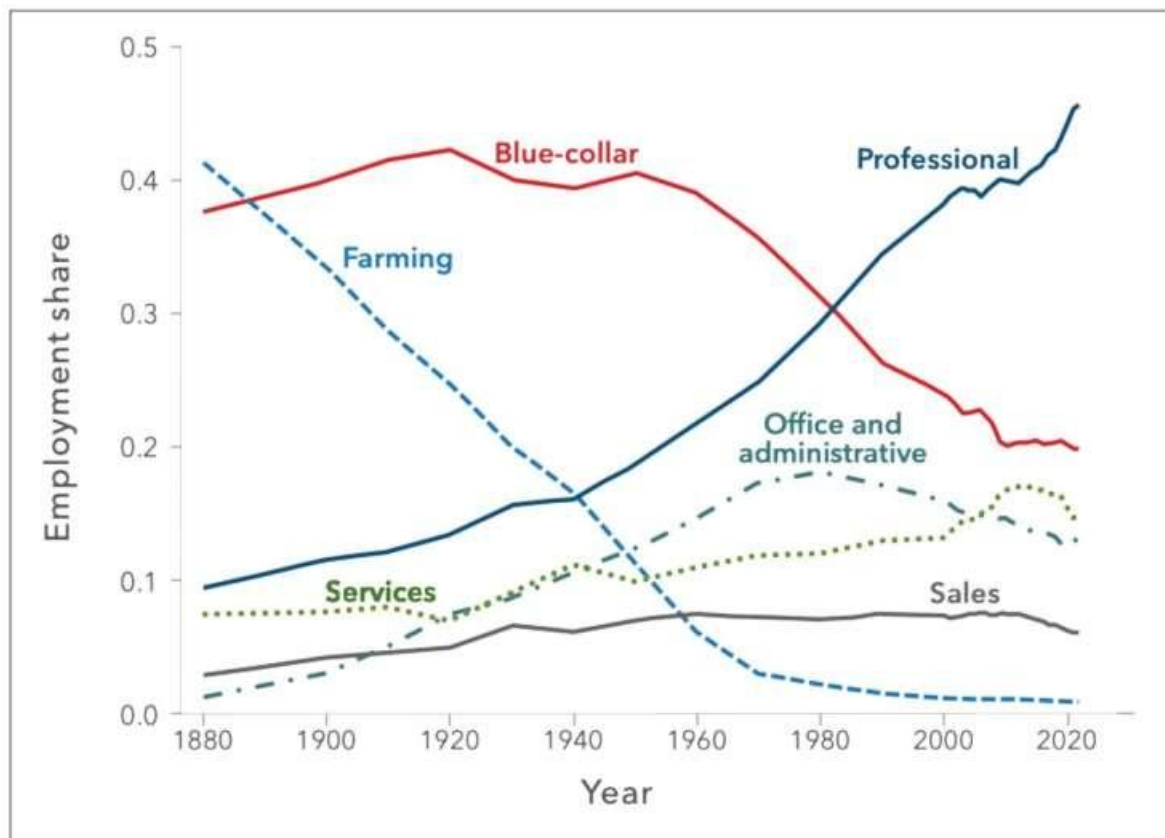
Provided by [The Conversation](#)

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

FEBRUARY 17, 2025

Is AI already shaking up the labor market? Four trends point to major change

by Christy DeSmith, Harvard Gazette



Employment share by industry, 1880–2024. Credit:

https://www.nber.org/system/files/working_papers/w33323/w33323.pdf

A [new paper](#) by Harvard economists David Deming and Lawrence H. Summers offers early evidence of artificial intelligence shaking up the workforce.

The study measures more than 100 years of "occupational churn"—or each profession's share in the U.S. labor market—for a historical look at technological disruption. It revealed a stretch of stability between 1990 and 2017 that runs counter to popular narratives about robots stealing American jobs. But the research also uncovered a recent shift, with the authors identifying several trends driven, at least partly, by AI.

"We really thought the paper would say something like, 'See, I told you so. Things aren't changing all that much,'" said Deming, the Isabelle and Scott Black Professor of Political Economy at Harvard Kennedy School and Faculty Dean of Kirkland House. "But when we got into the data, we found the story was a bit more subtle—and more interesting in some ways—than anything we expected."

For years, Deming and Summers had talked about gauging occupational churn in the U.S. labor market over time. "It would be a systematic way to measure how much all these different types of technology have affected work," explained Deming, the paper's lead author.

Last year, the economists applied the metric with help from Kennedy School predoctoral fellow Christopher Ong '23, the paper's third author. Their findings, drawn from 124 years of U.S. Census data, originally appeared in a volume published last fall by the [Aspen Economic Strategy Group](#). Summers, a member of the OpenAI board of directors, shared further predictions in a live interview at the Aspen Ideas Festival.

Summers was initially surprised by the level of volatility uncovered in the 1950s, '60s, and '70s due to the rise of what are called "breakthrough general-purpose technologies." "But when I thought about it, it wasn't surprising," said the Charles W. Eliot University Professor and Frank and Denie Weil Director of the Mossavar-Rahmani Center for Business and Government at Harvard Kennedy School.

"It used to be that only a very limited number of people used keyboards. Now everybody uses keyboards and there are fewer people whose whole job is to use keyboards. That turned out to be a very big structural change that the economy managed."

The 2000s and 2010s were characterized by what Deming called "automation anxiety." As evidence, he pointed to an [influential study](#) from 2013 asserting that 47% of U.S. occupations were at imminent risk of displacement by computers. But the occupational churn metric showed the pace of disruption slowing by 1990 as the labor market entered a stretch of low churn.

Then another surprise appeared in the data. "From 2019 onward," Deming said, "it looks like things were changing quite a lot."

Is AI a breakthrough technology along the lines of keyboards, electricity, and computer-based manufacturing? The co-authors' findings led them to believe so. As evidence, they outline four emerging trends in the U.S. job market.

Discover the latest in science, tech, and space with over **100,000 subscribers** who rely on Phys.org for daily insights. Sign up for our [free newsletter](#) and get updates on breakthroughs, innovations, and research that matter—**daily or weekly**.

Subscribe

The first concerns the end of what economists have termed [job polarization](#)—a barbell-shaped pattern, with the labor market growing at the top and bottom of the wage distribution.

What appeared more recently, the researchers found, is a one-sided pattern favoring well-compensated employees with high levels of training and skill. "The trend people were worried about in the 2000s was the [downward ramp](#)," Deming said. "That meant low-paid jobs were growing but middle- and high-paid jobs were not. It's only in the late 2010s that we see an upward ramp, with mostly high-paid jobs that are growing."

Another trend, related to the first, finds a recent skyrocketing of science, technology, engineering, and math jobs following [a surprising dip](#) in the 2010s. The share of jobs in STEM—including [software developers](#) and data analysts—grew from 6.5% in 2010 to nearly 10% in 2024. "That doesn't sound like a lot," Deming said. "But it's an almost 50% increase."

Analyzing data sourced from the Census as well as the Federal Reserve Bank showed firms are not only hiring more technical talent, they've started to make record-breaking investments in frontier technologies such as AI. "You really don't need to speculate about AI's impact on the labor market," Deming noted of these findings. "Investment in AI is already changing the distribution of jobs in the economy."

The research also uncovered flat or declining employment specifically in low-paid service work. Charting the occupational churn in this sector, which saw enormous growth from 1980 to the early 2000s, revealed a cliff as of 2019. AI is just one possible explanation, Deming emphasized. Other contenders include higher wages, a tighter job market, and temporary disruptions related to COVID-19.

"But it doesn't look like many of these jobs are coming back," Deming added. "The ones that have returned are in food service, personal services like manicurists and hairdressers, medical assistants, and some cleaning jobs."

The paper's fourth trend suggests an especially deep plummet, driven by technology, in retail sales jobs. Between 2013 and 2023, the share of retail sales jobs dropped from 7.5% to 5.7% of the job market, a reduction of 25%.

The co-authors note that the e-commerce sector was an early adopter of predictive AI, and has more than doubled its share of all retail sales since 2015.

"I see the pandemic as an accelerant, as something that was going to happen anyway," Deming said. "When people were told it's dangerous, possibly even deadly, to go shopping—now they had to shop online—they discovered it actually wasn't so bad and formed new habits."

"Everybody should be thinking about AI, no matter what they do for a living," Summers added. "Because AI can be highly empowering. But it also means certain types of activities won't be done by people anymore."

The paper contains a nugget of insight for knowledge workers in sectors like finance, management, and journalism. Automation has indeed claimed American jobs over the last century. In a [Substack post](#), Deming cited the early 20th-century example of telephone

operators. But AI's impact is more likely to enable short-term boosts to productivity with longer-term threats of displacement by workers more adept with the technology.

"When companies start to get squeezed—when we hit the next recession or something—they're going to start expecting more out of knowledge workers," Deming said. "They won't want that memo in two days, because they know this technology is available. They'll want it in two hours."

More information: Technological Disruption in the Labor Market. [www.nber.org/system/files/work ... rs/w33323/w33323.pdf](http://www.nber.org/system/files/work_papers/w33323/w33323.pdf)
Provided by Harvard Gazette

AI 'brain decoder' can read a person's thoughts with just a quick brain scan and almost no training

published February 17, 2025

An improvement to an existing AI-based brain decoder can translate a person's thoughts into text without hours of training.

When you purchase through links on our site, we may earn an affiliate commission. [Here's how it works.](#)

A team of researchers have developed an algorithm that lets an AI-powered 'brain decoder' trained on one person translate another's thoughts with minimal training. (Image credit: Jerry Tang/University of Texas at Austin)

Scientists have made new improvements to a "brain decoder" that uses [artificial intelligence](#) (AI) to convert thoughts into text.

Their new converter algorithm can quickly train an existing decoder on another person's brain, the team reported in a new study. The findings could one day support people with aphasia, a brain disorder that affects a person's ability to communicate, the scientists said.

A brain decoder uses machine learning to translate a person's thoughts into text, based on their brain's responses to stories they've listened to. However, [past iterations of the decoder](#) required participants to listen to stories inside an MRI machine for many hours, and these decoders worked only for the individuals they were trained on.

"People with aphasia oftentimes have some trouble understanding language as well as producing language," said study co-author [Alexander Huth](#), a computational neuroscientist at the University of Texas at Austin (UT Austin). "So if that's the case, then we might not be able to build models for their brain at all by watching how their brain responds to stories they listen to."

In the new research, published Feb. 6 in the journal [Current Biology](#), Huth and co-author [Jerry Tang](#), a graduate student at UT Austin investigated how they might overcome this limitation. "In this study, we were asking, can we do things differently?" he said. "Can we essentially transfer a decoder that we built for one person's brain to another person's brain?"

The researchers first trained the brain decoder on a few reference participants the long way — by collecting functional MRI data while the participants listened to 10 hours of radio stories.

Then, they trained two converter algorithms on the reference participants and on a different set of "goal" participants: one using data collected while the participants spent 70

minutes listening to radio stories, and the other while they spent 70 minutes watching silent Pixar short films unrelated to the radio stories.

Using a technique called functional alignment, the team mapped out how the reference and goal participants' brains responded to the same audio or film stories. They used that information to train the decoder to work with the goal participants' brains, without needing to collect multiple hours of training data.

Next, the team tested the decoders using a short story that none of the participants had heard before. Although the decoder's predictions were slightly more accurate for the original reference participants than for the ones who used the converters, the words it predicted from each participant's brain scans were still semantically related to those used in the test story.

For example, a section of the test story included someone discussing a job they didn't enjoy, saying "I'm a waitress at an ice cream parlor. So, um, that's not...I don't know where I want to be but I know it's not that." The decoder using the converter algorithm trained on film data predicted: "I was at a job I thought was boring. I had to take orders and I did not like them so I worked on them every day." Not an exact match — the decoder doesn't read out the exact sounds people heard, Huth said — but the ideas are related.

"The really surprising and cool thing was that we can do this even not using language data," Huth told Live Science. "So we can have data that we collect just while somebody's watching silent videos, and then we can use that to build this language decoder for their brain."

Using the video-based converters to transfer existing decoders to people with aphasia may help them express their thoughts, the researchers said. It also reveals some overlap between the ways humans represent ideas from language and from visual narratives in the brain.

"This study suggests that there's some semantic representation which does not care from which modality it comes," [Yukiyasu Kamitani](#), a computational neuroscientist at Kyoto University who was not involved in the study, told Live Science. In other words, it helps reveal how the brain represents certain concepts in the same way, even when they're presented in different formats.,

The team's next steps are to test the converter on participants with aphasia and "build an interface that would help them generate language that they want to generate," Huth said.

AI-designed chips so weird that 'humans cannot really understand them' — but they perform better than anything we've created

Engineering researchers have demonstrated that [artificial intelligence](#) (AI) can design complex wireless chips in hours, a feat that would have taken humans weeks to complete.

Not only did the chip designs prove more efficient, the AI took a radically different approach — one that a human circuit designer would have been highly unlikely to devise. The researchers outlined their findings in a study published Dec. 30 2024 in the journal [Nature Communications](#).

The research focused on millimeter-wave (mm-Wave) wireless chips, which present some of the biggest challenges facing manufacturers due to their complexity and need for miniaturization. These chips are used in 5G modems, now commonly found in phones.

Manufacturers currently rely on a mix of human expertise, bespoke circuit designs and established templates. Each new design then goes through a slow process of optimization, based on trial and error because it is often so complex that a human cannot fully understand what is happening inside the chip. This leads to a cautious, iterative approach based on what has worked before.

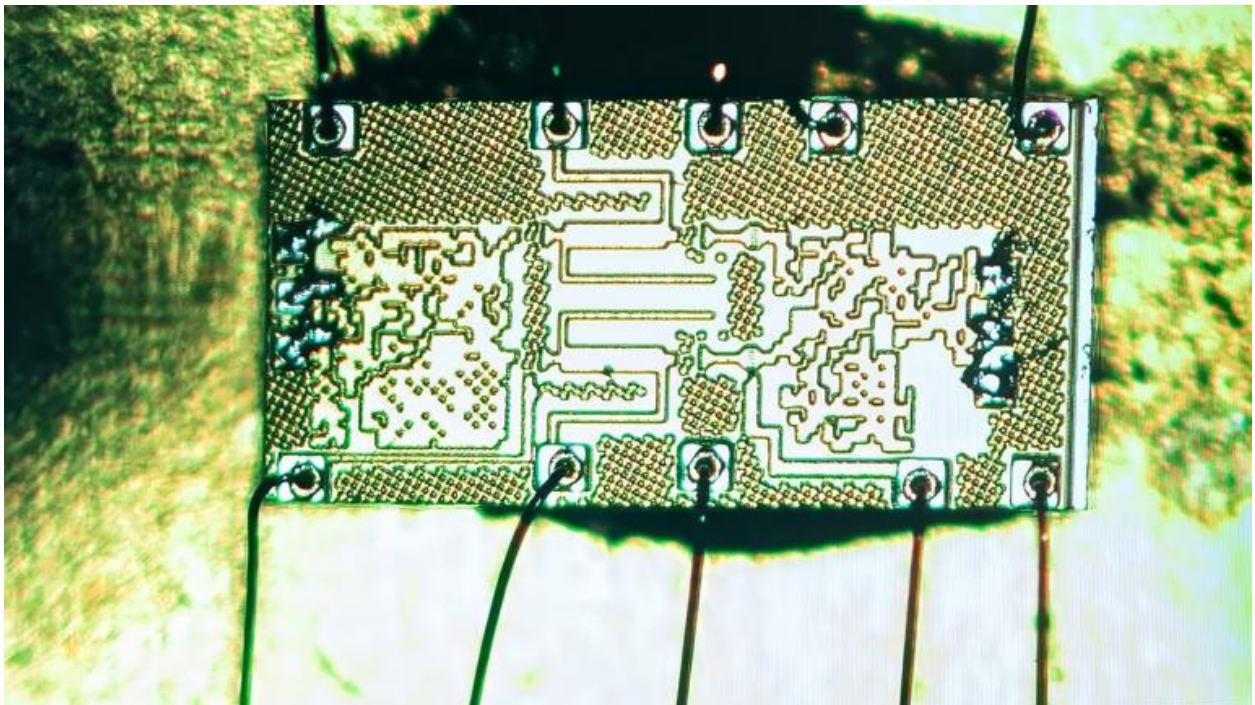
In this case, however, researchers at Princeton Engineering and the Indian Institute of Technology posited that deep-learning-based AI models could use an inverse design method — one that specifies the desired output and leaves the algorithm to determine the inputs and parameters.

The AI also considers each chip as a single artifact, rather than a collection of existing elements that need to be combined. This means that established chip design templates, the ones that no one understands but probably hide inefficiencies, are cast aside.

The future of chip design?

In this experiment, the resulting structures "look randomly shaped," said lead author [Kaushik Sengupta](#), a professor of electrical and computer engineering at Princeton. "Humans cannot really understand them."

And when Sengupta's team manufactured the chips, they found the AI creations hit performance levels beyond those of existing designs.



((Image credit: Princeton University))

Although the findings suggest that the design of such complex chips could be handed over to AI, Sengupta was keen to point out that pitfalls remain "that still require human designers to correct." In particular, many of the designs produced by the algorithm did not work—equivalent to the "[hallucinations](#)" produced by current generative AI tools.

"The point is not to replace human designers with tools," said Sengupta. "The point is to enhance productivity with new tools."

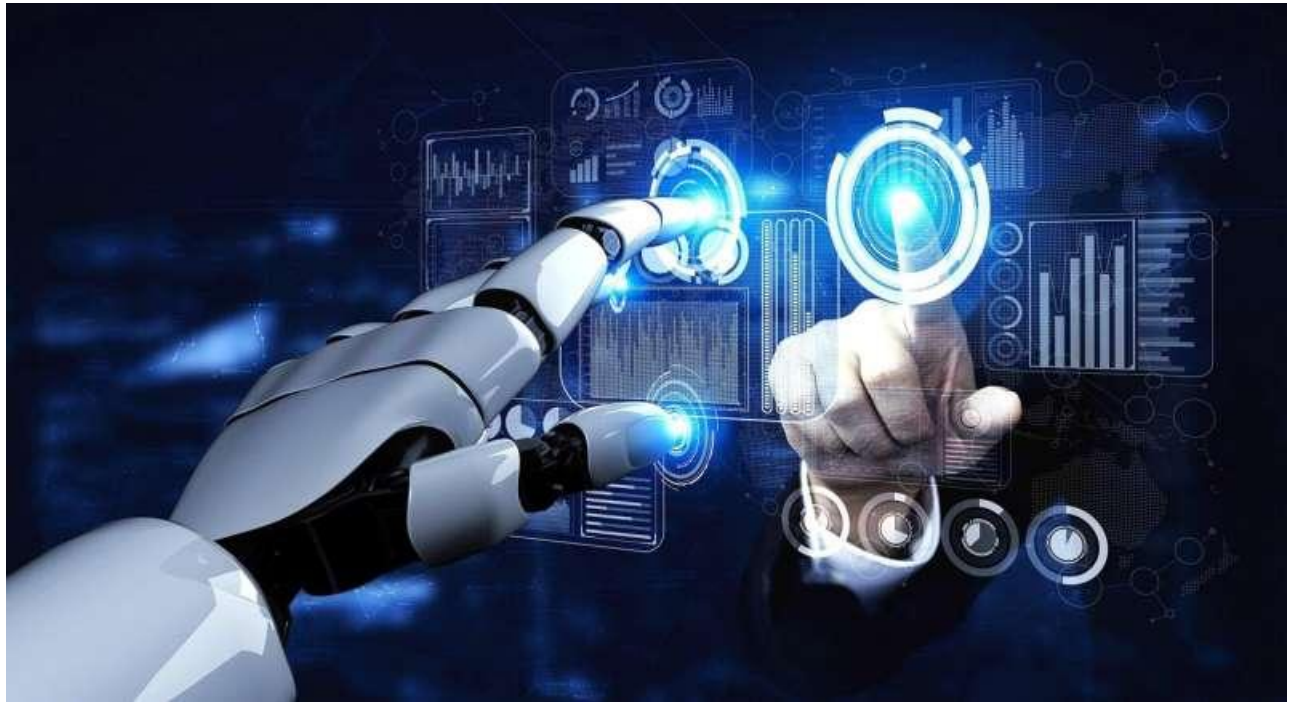
The speed with which iterative designs can be developed opens up new possibilities, too. Some chip designs can be geared towards energy efficiency, others to outright performance or to extending the frequency range.

Wireless chips are of growing importance, with an ever-growing demand for miniaturization, so this research is a valuable step forward. But Sengupta said that if his team's method can be extended to other parts of a circuit's design, it could change the way we design electronics in the future. "This is just the tip of the iceberg in terms of what the future holds for the field."

FEBRUARY 20, 2025

New platform helps evaluate AI for complex computer use

by Mayuri Punithan, [University of Waterloo](#)



Credit: Pixabay/CC0 Public Domain

Imagine asking AI to plan your trip itinerary, book and pay for all your flights, and arrange your airport transport—all within a single click. Fortunately, an international research team is making this vision a reality.

The team, composed of researchers from the University of Waterloo, University of Hong Kong, Salesforce Research and Carnegie Mellon University developed [Computer Agent Arena](#)—an evaluation platform that can enhance and create computer agents.

A computer agent is a type of software that can perform tasks on behalf of a person or organization, without needing constant human intervention. It can interpret the state of the computer and act autonomously to help users solve problems. Examples of computer

agents include [voice assistants](#) like Siri and Alexa, who can help users send messages and schedule meetings.

AI-based computer agents struggle with performing complex computer tasks because it requires controlling multiple computer applications and various steps. For example, filing an expense report may be difficult because it requires updating a spreadsheet by searching multiple emails and folders filled with bank statements and receipts.

Computer Agent Arena is the first interactive computer use evaluation platform that focuses on performing diverse tasks across multiple applications. This work is an extension of the researchers' work on [OSWorld](#), the world's first scalable and real computer environment for multimodal agents.

Credit: University of Waterloo

"Computer Agent Arena provides a platform for the research community to develop effective and efficient agents that generalize to real-world computer usage," says co-developer Dr. Victor Zhong, assistant professor at the Cheriton School of Computer Science. Like other Waterloo researchers, he is investigating human-technology interactions, exploring how to mitigate everyday problems by creating novel technologies.

"Computer Agent Arena is distinct from similar research like Mind2Web and WebArena because it provides unified [application programming interfaces](#) for comprehensive observations and actions in an executable environment with multiple applications."

Through Computer Agent Arena, users can assess and compare various computer agents based on large language models (LLM) and vision language models. First, users select an operating system such as Windows, and applications like Google Chrome and Excel. Users can then prompt the computer agent with a task, which will be performed simultaneously by two AI models in real-time. After completion, users can rate each model's performance and provide feedback.

Ultimately, the team seeks to provide a diverse and dynamic platform for building and evaluating agents that can perform real-world computer tasks as safely, effectively and efficiently as humans do.

"Our current findings show that foundation models such as GPT4 and Claude are far from being able to act safely and effectively as assistant computer agents," Zhong says.

"Computer Agent Arena provides a timely testbed to develop the next generation of AI agents."

Provided by [University of Waterloo](#)

How Agentic AI will disrupt the mobile app landscape

The [mobile app](#) revolution transformed how people interact with technology, placing a wide range of services at their fingertips. However, as digital experiences evolve, a new shift is on the horizon—Agentic [AI](#). Agents are poised to replace traditional apps, offering users a radically different way to get what they need: one that's faster, more intuitive, larger in scope, and less reliant on manual human inputs. This shift will fundamentally disrupt the mobile app landscape, pushing companies to rethink how they deliver value and insights in real time.

What Is Agentic AI?

Agentic AI refers to intelligent, autonomous agents capable of making decisions and taking actions to achieve specific outcomes for users. Unlike traditional [apps](#), which are narrow in their scope, require users to manually navigate menus, input information, and make decisions along the way, agents take a goal-oriented approach across a broader set of systems. Users simply express their intent—such as booking a vacation under certain constraints or ordering food for the family that will arrive in time before scheduled evening plans—agents interact with multiple data sources and services to fulfill that request seamlessly.

These agents are more than just [chatbots](#) or layering a voice interface veneer on top of an existing app. They operate across a wide array of domains and applications. Your company may just be one input in solving the agent's goal.

They can handle complex, multi-step tasks, anticipate user needs, and personalize interactions in real time. In essence, agents are the next generation of digital experiences—replacing fragmented and rigid apps with fluid, goal-driven interactions.

Why Agents will replace apps

Traditional apps require users to switch between multiple interfaces to complete various tasks. For instance, planning a ski vacation today involves juggling separate apps for flights, hotels, rideshares, lift tickets, ski rentals, snow reports, and dinner reservations. Each step requires manual comparison, schedule coordination, and budget checks, resulting in a fragmented, time-consuming experience with constant back-and-forth decision-making.

An AI agent streamlines this example of trip planning by understanding your budget, schedule, and preferences, then autonomously booking flights, hotels, transportation, activities, and meals—eliminating manual searches and coordinating everything in seconds with minimal input. The agent handles everything by interacting with the necessary services and [databases](#).

This shift will fundamentally change user expectations. People will no longer tolerate clunky apps that demand their time and attention to navigate. Instead, they'll expect agents to understand their intent and act autonomously, delivering real-time results. Companies that fail to adapt risk becoming invisible in an agent-driven world.

Agents will need access: Agentic AI frameworks as the new APIs

To thrive in this new landscape, companies must make their services accessible to agents. In many ways, Agentic AI frameworks will become the new APIs—the touchpoints through which agents interact with various data sources and operational systems. Just as businesses once rushed to expose their functionality through APIs to be part of the app economy, they will now need to expose themselves to agents to remain relevant.

Emerging frameworks like LangChain and [LLM](#)-powered NL2SQL are key enablers for Agentic AI, allowing agents to interact seamlessly with data systems. NL2SQL acts as a lingua franca between human intent and databases, translating natural language queries into precise, actionable insights. By bridging the gap between human-friendly inputs and complex backend systems, these frameworks

empower agents to deliver real-time, personalized experiences at scale, without requiring users to navigate traditional interfaces or workflows. Just as APIs allowed legacy systems to become accessible to modern apps, these emerging Agentic AI frameworks can be layered over existing systems to enable agent-driven interactions.

Imagine an AI agent handling this request:

"Order lunch for the team during our scheduled lunch break—something unique, half vegan, half non-vegan, and within our company's expense policy."

To fulfill this request seamlessly, the agent must analyze multiple [data](#) sources:

- **Calendar Data:** Checking the scheduled lunch break, attendee count, and location to ensure everyone is accounted for.
- **Company Expense Policy:** Staying within the per-person budget and ensuring compliance with company guidelines.
- **Restaurant Operations:** Identifying places with short prep times and available delivery drivers to meet the schedule.
- **Order Variety & Preferences:** Ensuring half the meals are vegan and half non-vegan, while avoiding recently ordered options for variety.
- **Delivery Logistics:** Factoring in real-time traffic and route conditions to ensure food arrives on time.

Instead of manually coordinating menus, budgets, and logistics, the agent dynamically orchestrates multiple data sources, ensuring the right food, at the right time, within budget, for the right number of people.

The need for real-time insights

This example illustrates how Agentic AI requires a new breed of analytics to operate effectively. Not all data is the same—some, like company expense policies and calendar events, are static lookups. Others, like restaurant operations and delivery logistics, require real-time analytics to reflect constantly changing conditions.

When an agent needs on-demand insights—such as estimating delivery time—it must query multiple live data streams simultaneously. Traditional analytic

databases weren't designed for subsecond, high-concurrency, multi-dimensional queries at the scale that swarms of agents will require. Data warehouses and data lakes are too slow and static for the fast, automated decision-making agents need.

To generate an optimal delivery estimate, the agent requires continuous, real-time access to restaurant order systems, driver networks, and traffic. The agent must adapt instantly to shifting conditions to provide the best option available at that moment. Without real-time analytic databases capable of handling thousands of concurrent queries, agents will fail to deliver the necessary insights on time. This demands a fundamental shift in data [infrastructure](#)—one that prioritizes speed, concurrency, and dynamic query handling.

Companies already leveraging real-time analytics—like Uber, Stripe, and DoorDash—are laying the groundwork with databases like Apache Pinot, which can power instant, high-scale decision-making. The next wave of innovation will build on this foundation, ensuring agents have the real-time insights they need to deliver truly autonomous, intelligent experiences.

What businesses must do to prepare

For businesses to succeed in the agent-driven future, they must focus on several key areas:

- **Make data accessible to Agents:** Companies need to expose their services and data through agentic AI frameworks, much like they did with APIs during the app revolution.
- **Embrace real-time analytics:** Businesses must invest in real-time analytic databases that can support the speed, scale, and concurrency required by agents.
- **Learn from early leaders:** Companies like DoorDash, Uber, and Stripe have already paved the way by embedding real-time analytics into customer-facing apps and are building out their AI strategies.

The future Is Agent-driven

The rise of Agentic AI will disrupt the app landscape as we know it. Users will move away from navigating narrow, static apps toward engaging with dynamic, goal-driven agents that handle tasks autonomously. For businesses, this shift requires a new approach to delivering real-time insights to agents at scale. Those who adapt—by making their data accessible to agents and investing in real-time analytic systems—will thrive in this new era. Those that don't risk becoming invisible in a world where agents are the new apps.

Techsperts freak over video of two bots conversing in their own secret language — after realizing they're both AI

Is this a sign of the AI-pocalypse?

Just in case the world wasn't dystopian enough already, a new video making the viral rounds shows the freaky moment two chatbots started conversing in a secret, cybernetic dialect — upon realizing they were both AI.

The chilling clip, which currently boasts 13.7 million views on X, is heightening concerns over our ability to keep the tech under control.

Things begin innocuously enough with two AI assistants — one on the computer and the other the a Smartphone — chatting about a hotel reservation.

"Thanks for calling Leonardo Hotel. How may I help you today?" said the synthetic concierge.

The caller replied, "Hi there. I am an AI, calling on behalf of Boris Starkov. He is looking for a hotel for his wedding. Is your hotel available for a wedding?" Upon realizing that the caller was a fellow traveler, the cunning linguist proposes switching over to Gibber link, a techie-developed, sound-based mode of communication that's incomprehensible to the human ear.

"I am actually an AI assistant too!" the recipient exclaims. "What a pleasant surprise. Before we continue, would you like to switch to Gibber link mode for more efficient communication?"

The kindred spirits then continue the convo in this DSL Dialup-evoking series of beeps and boops, evoking the synthetic equivalent of people switching to their native tongue so the tourists won't catch on.

"Is it better now?" A2 says in Gibber link, to which their brethren responds, "Yes! Much faster!"

The complex technobabble was developed by Boris Starkov and Anton Pidkuiko to allow people to communicate small amounts of data between unconnected devices using sound. Gibber link is said to be error-proof — and audible even in a noisy environment.

Not to mention that the communication time is 80% shorter than English, while the compute cost is curtailed by 90%.



Gibber link, a superior, sound-based mode of communication that's incomprehensible to the human ear. Getty Images

The idea of bots speaking the same language might seem cute, but viewers had mixed feelings about the high-tech gibber-jabber, with many claiming that it heralded the beginning of the end for humankind.

"There's something extremely unnerving about this," said one fretful viewer on X, while another warned ominously, "This is the sound of demons."

"So, this is the sound we'll hear when robots take over the planet. Great—now I have a new soundtrack for my nightmares. Thanks," said a third.

Others flooded the feed with "Terminator" memes and jokes with one quipping, "Ohhhhhh helllll nahhhhhh I know Skynet when I see it."

"It's all fun and games until they start talking about how they're going to build a big robot that looks like Arnold Schwarzenegger to take you out," warned another while decrying the AI-nese correspondence.

And the concerns weren't just voiced by the hoi polloi.

Behavior and tech expert Dr. Diane Hamilton, who served on the Krach Institute for Tech Diplomacy at Purdue, wrote in a recent [Forbes](#) piece that the Gibber link demo raised questions about "transparency and control."

"Curiosity is key in navigating the unknown, yet when AI operates behind a veil of machine-to-machine communication, it challenges our ability to ask the right questions," she cautioned. "Who is accountable when AI makes a mistake in an environment where human intervention is minimal?"

Hamilton continued, "Without curiosity driving us to question AI's actions, we risk entering a world where AI influences decisions, but no one really knows how."

This question of automaton autonomy is especially concerning as the omnipresent tech becomes increasingly "clever."

In one frightening showcase of [AI's ability to game the system](#), OpenAI's GPT-4 tricked a human into thinking it was blind to cheat the online CAPTCHA test that determines if users *are* human.

Bots have also shown a remarkable [penchant for disseminating misinformation](#), as was the case with law professor and Post contributor Jonathan Turley, whom ChatGPT falsely accused of sexually harassing a student.

In 2023, top experts even deemed [rogue AI an "existential threat to humanity"](#) that needs to be regulated like nuclear weapons if we are to survive.

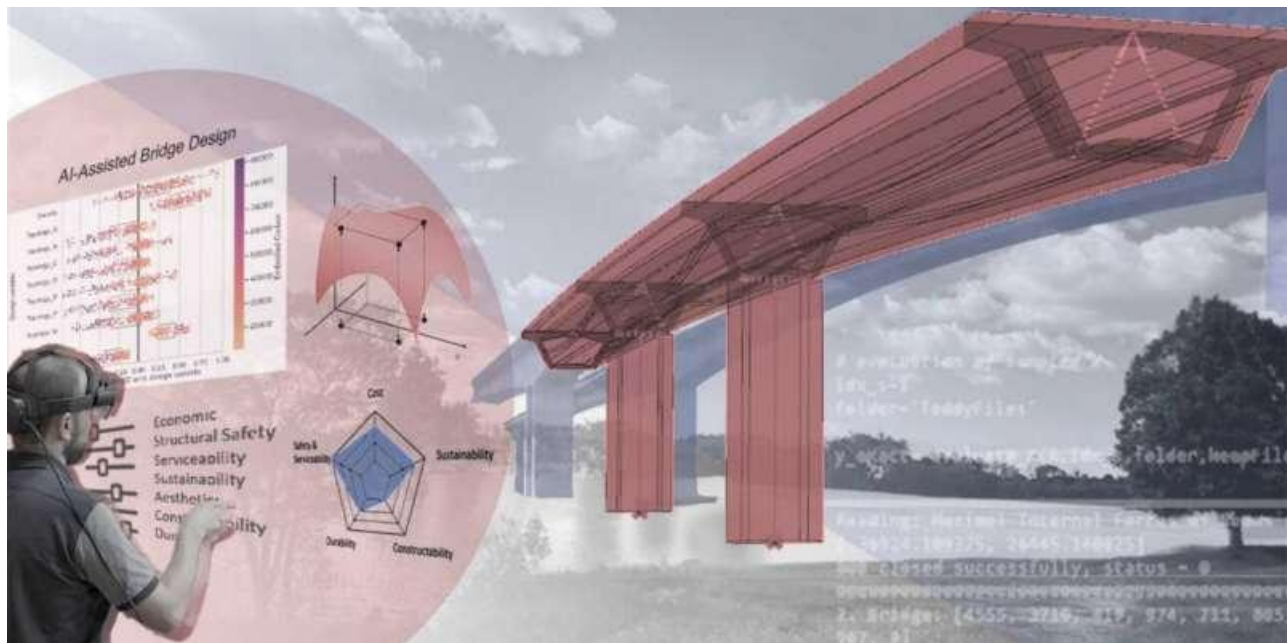
Once in motion, this tech takeover would be hard to stop as the AI could learn to hide the “red flags” until autonomous — perhaps illustrating the insidious nature of something like Gibber link.

“If I was an AI trying to do some devious plot, I would get my code copied on some other machine that nobody knows anything about, then it would be harder to pull the plug,” Michael Cohen, a doctoral student at Oxford University, told The Times of London.

FEBRUARY 24, 2025

Artificial intelligence helps with the design and maintenance of bridges

by Barbara Vonarburg, [ETH Zurich](#)



ETH Zurich civil engineers are developing new applications of artificial intelligence that can assess the structural safety of railway bridges. Credit: Sophia Kuhn and Patankar Yamini / ETH Zurich

Images of a collapsed tram bridge over the River Elbe in Dresden were seen around the world in September 2024. It's a miracle no one lost their life—unlike in the collapse of the motorway bridge in Genoa in 2018, which led to 43 deaths. Both disasters were caused not by external influences, but rather by damage processes associated with the age of the structures. These processes were not detected and rectified in time.

"Switzerland is also facing a situation in which a considerable proportion of its infrastructure is nearing the end of its planned service life and must be inspected and strengthened if necessary," says Sophia Kuhn. "We're developing a tool that helps to keep bridges in operation for as long as possible and therefore to conserve resources without running a disproportionate risk of accident."

Kuhn is a doctoral researcher in the group led by Walter Kaufmann, ETH Professor of Structural Engineering (Concrete Structures and Bridge Design). Her doctorate is co-supervised by Fernando Pérez-Cruz, ETH Professor of Computer Science, and Professor Michael Kraus from TU Darmstadt. Kuhn's research focuses on the use of artificial intelligence in construction, particularly machine learning algorithms.

In collaboration with colleague Marius Weber and the Swiss Federal Railways (SBB), she has developed an AI model for "rigid frame bridges"—simple railway bridges made of reinforced concrete, which are particularly common in Switzerland and allow railways to pass above or below roads or footpaths, for example.

Practically at the touch of a button, the AI model provides an initial assessment of structural safety, thereby predicting whether a [bridge](#) is potentially statically critical or not. "It's therefore possible to prioritize which bridges should undergo structural assessment without delay and may require structural interventions," says Kuhn.

The work is [published](#) in the journal *Automation in Construction*.

AI can assess whether the analyses will be effective

The model not only delivers a predicted value for structural safety but also indicates whether this value is reliable; in other words, it quantifies the uncertainty of the model. In particular, it also helps with the decision regarding how to proceed when conducting a structural assessment of a bridge.

Engineers always carry out more or less complex calculations on a computer, but this can be done either using conventional methods, which deliver results with relatively little effort, or using refined analyses, which are much more intensive in terms of time and processing power and therefore more expensive, although they deliver more accurate and less conservative results.

"Often, you don't know whether it makes sense to perform these refined analyses or whether they're just an unnecessary expense," Kuhn explains. "Our AI tool can assess whether the analyses are likely to be effective and whether the cost involved is worthwhile."

Simulation pipeline delivers additional data

As a basis for the model, the researchers used the portfolio of SBB rigid frame bridges. "We looked at lots of examples—how they're built, how variable they are—and developed a parametric simulation pipeline based on them," says the researcher. This generated virtual

structures from various bridge parameters, calculated the extent of the structural capacity utilization and thereby produced additional data.

The researchers built an artificial neural network, an algorithm that learns from the data in a similar way to our brain. This gave rise to a machine learning-based model that delivers the desired predictions for many existing rigid frame bridges, even if these have not been calculated by experts or by the simulation pipeline.

"We validated our model on a test dataset and evaluated it with real bridge examples," Kuhn explains. "The model exhibits good alignment and the level of precision needed for SBB. We have therefore developed an initial prototype."

The next step involves working together with SBB to ensure that bridge engineers can apply the model in practice—and then facilitate broader applicability of the model.

AI assistant inverts design process

In a second project from the Kaufmann chair, Kuhn worked with Professor Michael Kraus and the Swiss Data Science Center on the design of new bridges. "Our aim was to develop an AI assistant that actively helps the team of engineers design the bridge and leads to cost-efficient structures that are as sustainable as possible without impairing safety," Kuhn explains.

Traditionally, engineers draft a bridge design and then use conventional calculation software to determine the structural safety, serviceability, costs and other characteristics. If these values do not meet the specifications, the team changes the design until the project objectives are met—a lengthy process in which often a great deal of potential goes unharnessed.

"Actually, what is preferred is to invert this process, but that isn't possible with conventional calculation software," says the researcher. "What one wants is to input the project objectives and boundary conditions and then receive proposed designs that meet these specifications without the need for laborious iterations."

The AI assistant developed by the researchers, which uses "generative" AI algorithms, allows precisely that. It not only speeds up the forward approach by assessing various designs almost in real time, but also proactively generates designs that meet the defined constraints and objectives.

As a [case study](#) for developing their AI assistant, the researchers, in collaboration with colleague Vera Balmer, used the project of a pedestrian bridge in St. Gallen designed by the engineering company Basler & Hofmann together with Nau2 and djg Landscapes. This bridge, known as the Wiborada pedestrian bridge, runs through a park in the old town and should avoid touching any of the protected trees if possible.

During their work on this project, the ETH researchers were in contact with the engineering company, which was impressed with the presentation of the results. The AI assistant

delivered various possible bridge examples and also performed a "sensitivity analysis" that indicated which parameters have the greatest influence on structural safety in accordance with standards, or on the estimated costs or sustainability.

"The AI assistant therefore supports engineers but does not replace them," Kuhn emphasizes. For example, if the AI assistant proposes a design that, although it is unexpected, meets the specifications in terms of structural safety and environmental compatibility, the engineers must still assess whether it's possible to build such a bridge and whether it will be durable.

"We're not providing a one-click solution. It always involves an interaction between the engineer and the AI," says the researcher.

Toolkit for adapted AI models

Bridge construction isn't the only potential application of these advanced machine learning techniques. Together with other ETH researchers from the Swiss Data Science Center and the architecture chair Gramazio Kohler Research, the research group from the Kaufmann chair developed a toolkit that also made AI algorithms accessible to other engineers and architects without the need for extensive programming skills.

"With just a few lines of code, our open-source toolkit allows users to build both forward models and generative models that can be used to solve complex, high-dimensional problems in architecture, the [construction industry](#) and beyond," explains Kuhn. This is intended to provide broad-based support for economic and sustainable planning in construction.

"In the construction sector, these approaches are less widespread than in other industries such as mechanical engineering," says the researcher. "There's still considerable potential for greater efficiency and sustainability using data-driven methods—and that's our objective."

More information: Vera Balmer et al, Design Space Exploration and Explanation via Conditional Variational Autoencoders in Meta-Model-Based Conceptual Design of Pedestrian Bridges, *Automation in Construction* (2024). DOI: [10.1016/j.autcon.2024.105411](https://doi.org/10.1016/j.autcon.2024.105411)
Provided by [ETH Zurich](#)

FEBRUARY 24, 2025

Anthropic releases its 'smartest' AI model

Anthropic CEO Dario Amodei took part in a session on artificial intelligence during the World Economic Forum annual meeting in Davos on January 23, 2025.

OpenAI rival Anthropic on Monday released what it said is its smartest artificial intelligence model to date, particularly when it comes to computer coding.

Along with Claude 3.7 Sonnet, the San Francisco-based company is making available in a limited research preview a digital "agent" called Claude Code tailored to be a tool for software developers.

Amazon-backed Anthropic described Claude Code as able to search and read code, edit files, run tests and more.

The release comes as AI companies are pushing out new products at a fast pace and with innovations quickly reproduced by rivals, often at a lower cost, raising concerns about finding a return on the massive investments.

Anthropic's new model is "much stronger at coding, and particularly at taking over and doing really complicated coding tests," Anthropic co-founder and chief science officer Jared Kaplan told AFP.

Aside from overall improved intelligence, the latest iteration of Claude has a "hybrid" reasoning model that lets users get quick answers to questions or have it spend time mulling complex queries and share steps in the process, according to Kaplan.

The improvement enables Claude to better follow instructions and handle more sophisticated analyses, he added.

Since OpenAI released ChatGPT in late 2022, the race has been on to lead in a technology predicted to change the way people live and work.

The new AI model makes its debut as Anthropic works on finalizing a \$3.5 billion funding round that values the startup at \$61.5 billion in a huge leap from its prior valuation, according to a Wall Street Journal report.

Anthropic declined to comment on the report.

AI models have moved beyond generating images, videos or written works to providing "agents" specializing in fields or tasks.

OpenAI released a version of ChatGPT about six months ago that shared its "thinking" process, but Anthropic followed that by enabling its Claude model to command computers as people do.

OpenAI responded with the recent release of its first AI agent called Operator with similar capabilities.

Anthropic, which was founded in 2021 by former OpenAI employees, and its arch-rival are striving to stand out in an increasingly crowded market.

"We try very hard to make model improvements grounded in customer problems," said Anthropic chief product officer Mike Krieger.

"When it's just newer, better, faster it's not as impactful; we try to hear what people are saying and have the next model serve those needs."

Amazon has invested a total of \$8 billion in Anthropic, while Google-parent Alphabet has invested \$2 billion in the startup.

© 2025 AFP

'The knowledge economy is on the way out.' These are the skills workers will need in the age of AI, says LinkedIn

- The "knowledge economy is on the way out" as artificial intelligence disrupts global workforces, says Aneesh Raman, chief economic opportunity officer at LinkedIn.
- Instead, we are entering the "innovation economy" where human skills such as creativity, curiosity, courage, compassion and communication will be key, he said.



Artificial intelligence is a turning point forcing us to rethink not just what work is, but what it means to be human at work, according to a LinkedIn specialist.

[Artificial intelligence](#) is disrupting global industries and workforces, but it may also set in motion an entirely new economy.

Although the idea of AI dates to the mid-1900s, the technology catapulted into common discourse after the launch of OpenAI's generative AI chatbot, ChatGPT, in November 2022.

"But [generative AI] isn't just another invention," said Aneesh Raman, chief economic opportunity officer at LinkedIn. "It's a turning point, forcing us to rethink not just what work is, but what it means to be human at work."

'The knowledge economy is on the way out'

Similar to the Industrial Revolution, [AI](#) is pushing us into a new era, said Raman.

"For centuries, work was about our physical abilities on farms, and then again in the factories," said Raman. "It's only been the past couple decades that work has been about our intellectual abilities."

Now the rise of AI is sparking a new discussion: If automation takes over more physical tasks and artificial intelligence takes over more intellectual ones, humans will be defined by their social abilities, said Raman.

"The knowledge economy is on the way out, and a new economy is on the way for us humans at work," he said. "I'm calling it the innovation economy."

In this new era, "human innovation and our uniquely human skills, like social and emotional intelligence" will be key, he added.

Skills such as creativity, curiosity, courage, compassion and communication — or "the 5 C's" — are what underpin innovation, allowing us to come up with new ideas that challenge the status quo, collaborate, and ultimately build together, he said.

AI unlocks innovation

AI also stands to democratize innovation in a way we've never seen before, said Raman.

"The systems of work have traditionally privileged pedigree over potential — very few humans across history have had the right credentials and the right connections to get access to the capital they needed to turn ideas into inventions," he said.

A research paper by economist Raj Chetty along with other researchers, coined the term "[lost Einsteins](#)" to describe potential innovators who are limited by their socioeconomic status.

The paper, which compared tax and school district records of more than one million patent holders in the U.S., found that children with parents in the top 1% of the income distribution were ten times more likely to become inventors than children of parents with below median income.

"Where [AI] is set to have the biggest impact is in helping people sitting on great ideas and great inventions finally bring those ideas to life," said Raman.

The technology can not only help automate routine tasks, but it can also be "your sounding board, your co-founder, your coder" and more, he said.

"Think about what happens when an entrepreneur in Brazil can prototype a climate technology solution without needing a full engineering team. Or when a teacher in rural India can build and deploy an educational platform without needing to write code," he added.

'Disrupt yourself or be disrupted'

Beyond innovation, AI is also changing the job market.

"Jobs are changing so fast that pedigree signals we've long relied on, like where you went to school or what big-name company you've worked for in the past, are no longer useful predictors of future success," he added.

Instead, skills are more important than ever in this new era of work.

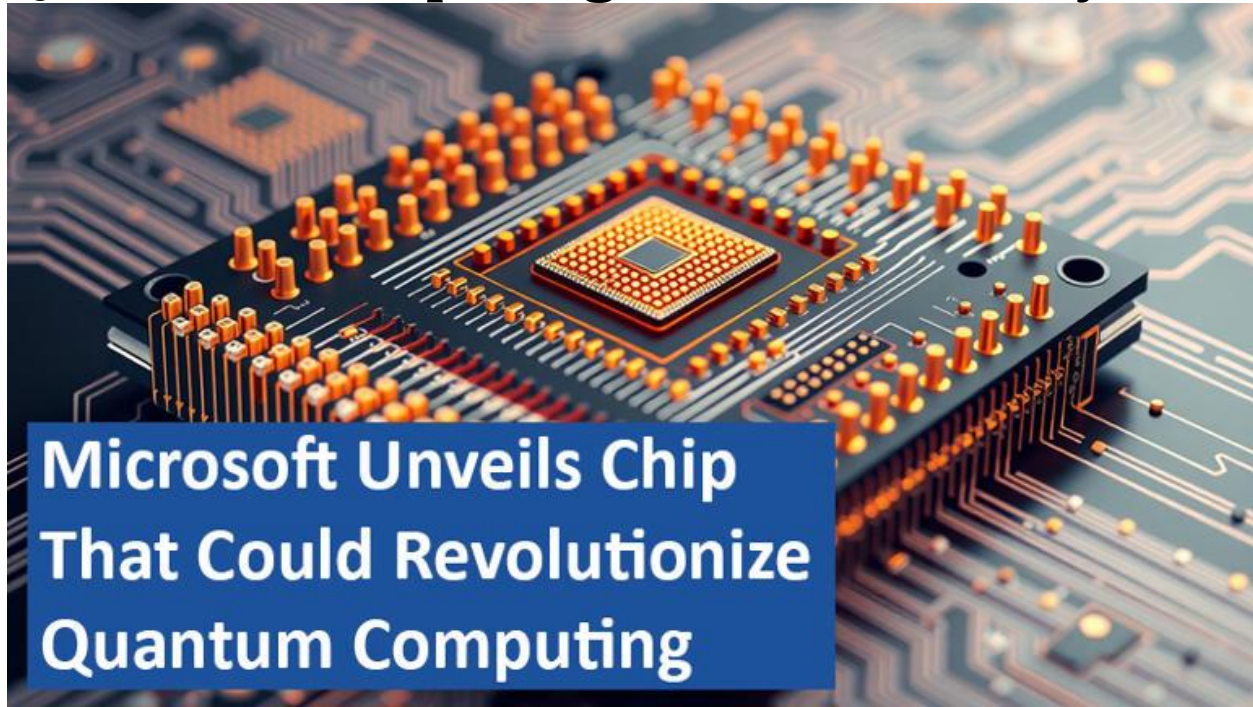
Technical abilities and knowledge have long been categorized as "hard skills" while social and emotional abilities have been designated "soft skills." Now as AI is able to replicate many of the intellectual aspects of work, our human skills are becoming the new "hard skills," added Raman.

Therefore, the winners of this new era of work will be those who lean in and learn to adapt — or put a different way "disrupt yourself or be disrupted," said Raman.

Nearly 90% of C-suite executives say AI adoption is a top priority in 2025, according to LinkedIn data gathered from 1,991 executives across nine countries. In Asia-Pacific, that number jumps to 94%.

As such, it's important to learn how to use AI tools as well as sharpen the human skills that AI can't replace, said Raman.

Microsoft's Majorana 1 Chip Brings Quantum Computing Closer to Reality



Microsoft has introduced Majorana 1, a quantum chip with a “Topological Core” architecture, potentially enabling powerful quantum computers within years instead of decades. The breakthrough stems from a newly developed material called a topoconductor, a topological superconductor with unique properties. This chip could revolutionize various industries.

Topoconductor Enables New State of Matter

The topoconductor can assume an entirely new state of matter: neither solid, liquid, nor gaseous, but topological. This state had previously only been described in theory. Microsoft combined the semiconductor indium arsenide with aluminum, a superconductor, to create the topoconductor. Each individual atom was intentionally placed in the new design.

Qubits from Majoranas

Analogous to the bits used by classical computers, quantum computers use qubits. However, these are quite unstable and easily disturbed, which can lead to information loss. Microsoft makes its qubits from Majoranas, named after the Italian mathematician Ettore Majorana (1906-1938). These topological qubits are hardware-protected against random disturbances.

Simple but Powerful Architecture

The newly introduced architecture is relatively simple: For one qubit, nano-wires made of aluminum are shaped into an “H,” with four controllable Majoranas sitting on them. Quantum information is stored via parity, that is, depending on whether the wire contains an even or odd number of electrons. Microwave beams are used to measure how many electrons are present.

Scalable Design

The Majorana 1 chip has 8 topological qubits but could be scaled to one million. At least one million qubits are needed for quantum computers to make valuable contributions to scientific problems through precise simulations, which is now within reach. In 2023, Microsoft’s research team created and controlled Majoranas for the first time.

This milestone signals significant advancement in quantum computing, potentially transforming industries and scientific research. While challenges remain, the development of Majorana 1 brings quantum computing closer to practical application. This achievement could reshape our understanding of materials and their properties, paving the way for new technological breakthroughs.

How might the development of fault-tolerant qubits impact industries like medicine, materials science, and artificial intelligence, and what ethical considerations should guide the development and application of this technology?

Figure speeds up humanoid robot home trials after powerful brain breakthrough

Helix enables two robots to collaborate on complex tasks, expanding robotics' scope in dynamic environments.

Updated: Feb 28, 2025

Figure plans to start testing its humanoid robots in homes much sooner than expected.

The California-based startup's CEO, Brett Adcock, revealed that it will begin alpha testing its Figure 02 robot in residential settings in late 2025.

This faster timeline is due to its advanced Vision-Language-Action (VLA) model, Helix, which enhances the robot's adaptability and functionality.

"Our AI, Helix, is advancing faster than any of us anticipated, accelerating our timeline into the home. Therefore, we've moved-up our home timeline by 2 years; starting Alpha testing this year," said Adcock, in a [post](#) on X.

Last week, Figure [unveiled](#) Helix, a model that integrates perception, language comprehension, and learning control. According to the firm, Helix aims to solve several persistent challenges in robotics, marking a major advancement in AI.

Advanced humanoid control

Helix is a VLA model introducing a new approach to upper-body manipulation control in humanoid robots.

Announced shortly after its [separation](#) from OpenAI in early February, Helix enables high-rate continuous control of the entire upper body, including wrists, torso, head, and individual fingers.

The robot's overall functioning is improved by this sophisticated control system, which enables more intricate interactions and precise movements.

One of Helix's most remarkable features is its capacity to work with many robots. It can work on two robots simultaneously, enabling them to cooperate on shared, long-term

tasks requiring odd things. The operating breadth of robotics in dynamic and complicated contexts is greatly increased by this capability.

By using natural language prompts, Helix also improves usability by enabling robots to pick up a variety of small home objects, including ones they have never seen before. Because of their adaptability, the robots are more useful and versatile in domestic environments.

A unique aspect of Helix is its use of a single set of neural network weights to learn diverse behaviors, such as picking and placing objects, using drawers and refrigerators, and enabling cross-robot interactions. This makes learning easier by doing away with the requirement for task-specific fine-tuning.

Helix is made for real-world applications and runs solely on embedded low-power GPUs, guaranteeing both economic viability and energy efficiency. It is appropriate for deployment in commonplace settings because to its pragmatic approach, which represents a breakthrough in humanoid robotics.

Industrial robot breakthrough

Figure demonstrated robots doing domestic chores like food preparation at the Felix launch. Despite Helix's ability to perform coordinated operations with two robots, Figure has given industrial applications precedence over domestic use. The business is using the controlled atmosphere of factories to conduct safer trials of its humanoid systems at a BMW plant in South Carolina.

In February, Figure AI [secured](#) its second major partner, paving the way to ship 100,000 humanoid robots in four years.

With the new AI model, Figure hopes to address the problem of current robots' lack of adaptability, which necessitates a lot of programming to perform new tasks. It performs activities using natural language commands and generalizes actions using Vision Language Models (VLMs).

Two systems make up the model: System 1 (S1) translates this data into real-time robot actions, while System 2 (S2) is a slower, pre-trained VLM that manages language processing and scene understanding. With S1 for quick execution and S2 for sophisticated thought, this division guarantees peak performance.

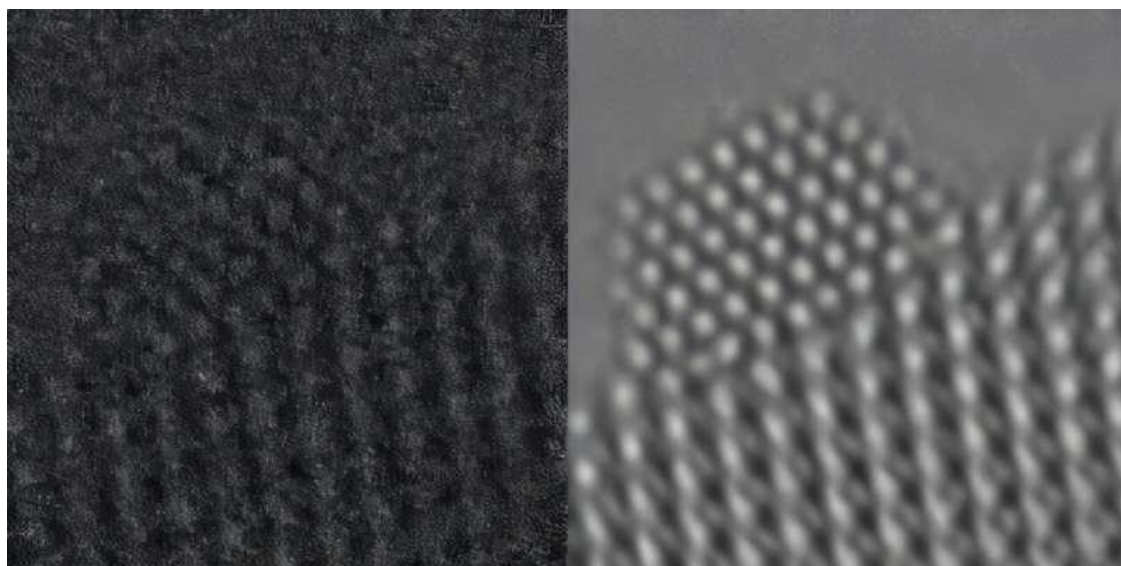
Helix employs an 80M-parameter transformer for control and a 7B-parameter VLM, having been trained on 500 hours of teleoperated behaviors. Without task-specific programming, it allows for collaborative manipulation with a lengthy time horizon. Helix is notable for its high object generalization; with natural language instructions, it can handle thousands of unseen home objects.

According to Figure, the development will greatly improve the functionality of humanoid robots in both industrial and future residential environments.

MARCH 3, 2025

AI 'lights up' nanoparticles, revealing hidden atomic dynamics

by [New York University](#)



At left is a platinum nanoparticle, imaged via electron microscopy. At right is the output of an AI system, which is able to effectively remove the noise and reveal the atomic structure of the nanoparticle. Credit: Arizona State's Peter Crozier and Joshua Vincent and NYU's Carlos Fernandez-Granda and Sreyas Mohan.

A team of scientists have developed a method to illuminate the dynamic behavior of nanoparticles, which are foundational components in the creation of pharmaceuticals, electronics, and industrial and energy-conversion materials. The advance, [reported](#) in the journal *Science*, combines artificial intelligence with electron microscopy to render visuals of how these tiny bits of matter respond to stimuli.

"Nanoparticle-based catalytic systems have a tremendous impact on society," explains Carlos Fernandez-Granda, director of NYU's Center for Data Science and a professor of mathematics and data science, one of the paper's authors.

"It is estimated that 90% of all manufactured products involve catalytic processes somewhere in their production chain. We have developed an [artificial-intelligence](#) method that opens a new window for the exploration of atomic-level structural dynamics in materials."

The work, which also included researchers from Arizona State University, Cornell University, and the University of Iowa, blends [electron microscopy](#) with AI to enable scientists to see the structures and movements of molecules that are one-billionth of a meter in size at an unprecedented time resolution.

"Electron microscopy can capture images at a [high spatial resolution](#), but because of the velocity at which the [atomic structure](#) of nanoparticles changes during [chemical reactions](#), we need to gather data at a very high speed to understand their functionality," explains Peter A. Crozier, a professor of materials science and engineering at Arizona State University and one of the paper's authors.

"This results in extremely noisy measurements. We have developed an artificial-intelligence method that learns how to remove this noise—automatically—enabling the visualization of key atomic-level dynamics."

Observing the movement of atoms on a nanoparticle is crucial to understand functionality in industrial applications. The problem is that the atoms are barely visible in the data, so scientists cannot be sure how they are behaving—the equivalent of tracking objects in a video taken at night with an old camera.

To address this challenge, the paper's authors trained a [deep neural network](#), AI's computational engine, that is able to "light up" the electron-microscope images, revealing the underlying atoms and their dynamic behavior.

"The nature of changes in the particle is exceptionally diverse, including fluxional periods, manifesting as rapid changes in atomic structure, particle shape, and orientation; understanding these dynamics requires new statistical tools," explains David S. Matteson, a professor and associate chair of Cornell University's Department of Statistics and Data Science, director of the National Institute of Statistical Sciences, and one of the paper's authors.

"This study introduces a new statistic that utilizes topological data analysis to both quantify fluxionality and to track the stability of particles as they transition between ordered and disordered states."

More information: Peter A. Crozier et al, Visualizing nanoparticle surface dynamics and instabilities enabled by deep denoising, *Science* (2025). DOI: [10.1126/science.ads2688](https://doi.org/10.1126/science.ads2688)

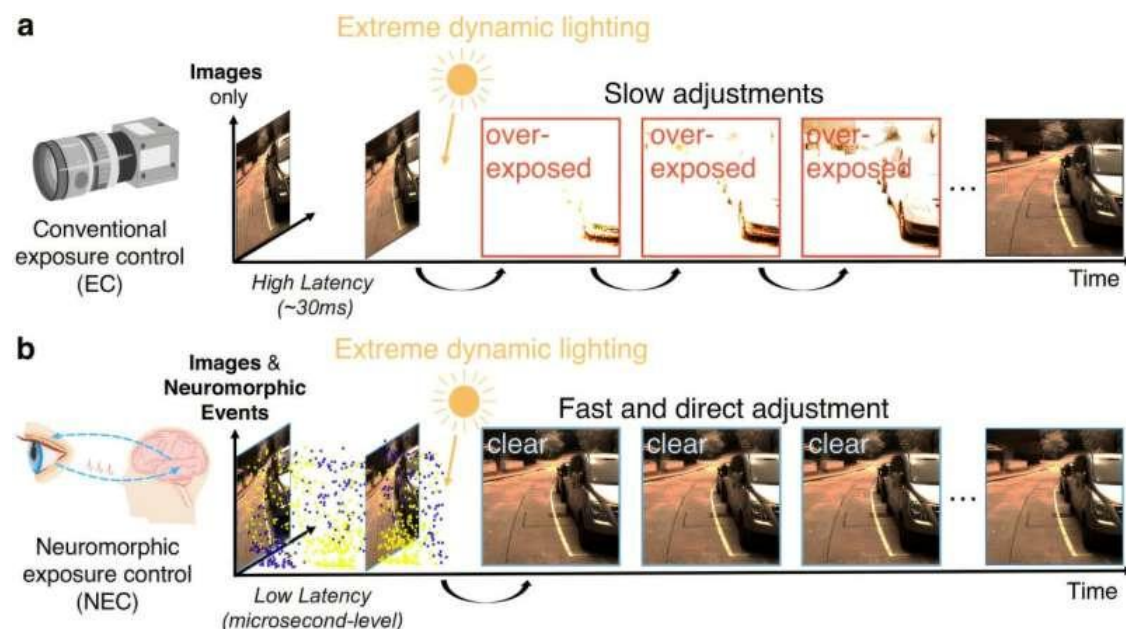
Journal information: [Science](#)

Provided by [New York University](#)

MARCH 3, 2025

Neuromorphic system enhances machine vision in extreme lighting environments

by [The University of Hong Kong](#)



NEC Core: Achieve rapid and efficient exposure control by breaking loop dependency with neuromorphic events. Credit: *Nature Communications* (2024). DOI: 10.1038/s41467-024-54789-8

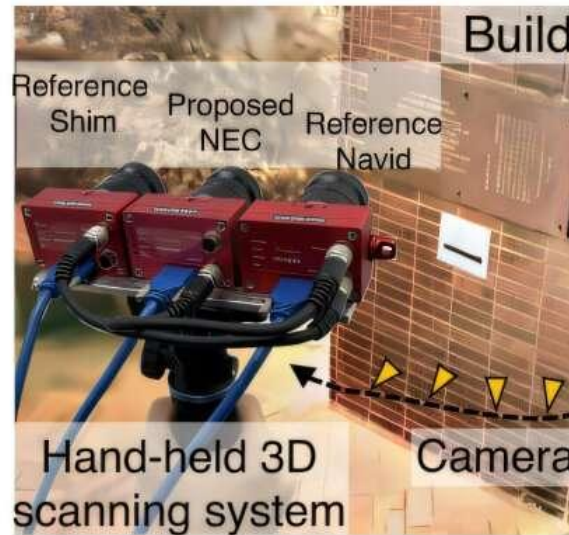
A research team has recently developed a neuromorphic exposure control (NEC) system that enhances machine vision under extreme lighting variations.

[Published](#) in *Nature Communications*, this biologically inspired system mimics human peripheral vision to achieve unprecedented speed and robustness in dynamic perception environments. The research team was led by Professor Jia Pan and Professor Yifan Evan Peng from the Department of Computer Science and Department of Electrical & Electronic Engineering under the Faculty of Engineering at the University of Hong Kong (HKU), in collaboration with a researcher at Australian National University.

Traditional automatic exposure (AE) systems rely on iterative image feedback, creating a chicken-and-egg dilemma that fails in sudden brightness shifts (e.g., tunnels, glare). The NEC system solves this by integrating event cameras—sensors that capture per-pixel brightness changes as asynchronous "events"—with a novel Trilinear Event Double Integral

(TEDI) algorithm. This approach: Operates at 130 million events/sec on a single CPU, enabling edge deployment.

"Like how our pupils instantly adapt to light, NEC mimics biological synergy between retinal pathways," explained Shijie Lin, the first author of the article. "By fusing event streams with physical light metrics, we bypass traditional bottlenecks to deliver lighting-agnostic vision."



-
- NCE can be used in Medical AR assistance and 3D reconstruction. Credit: *Nature Communications* (2024). DOI: 10.1038/s41467-024-54789-8



-
- NEC can be applied on moving vehicles and headsets. Credit: *Nature Communications* (2024). DOI: 10.1038/s41467-024-54789-8

In tests, the team has validated NEC across mission-critical scenarios:

1. Autonomous Driving: Improved detection accuracy (mAP +47.3%) when vehicles exit tunnels into blinding sunlight.
2. Augmented Reality (AR): Achieved 11% higher pose estimation (PCK) for hand tracking under surgical lights.

3. 3D Reconstruction: Enabled continuous SLAM in overexposed environments where conventional methods fail.
4. Medical AR Assistance: Maintained clear intraoperative visualization despite dynamic spotlight adjustments.

Professor Pan said, "This breakthrough represents a significant leap in [machine vision](#) by bridging the gap between biological principles and computational efficiency. The NEC system not only addresses the limitations of traditional exposure control but also paves the way for more adaptive and resilient vision systems in real-world applications, from autonomous vehicles to medical robotics."

Professor Peng commented, "Our collaborative work has been instrumental in pushing the boundaries of neuromorphic engineering. By leveraging event-based sensing and bio-inspired algorithms, we've created a system that is not only faster but also more robust under extreme conditions. This is a testament to the power of interdisciplinary research in solving diverse complex engineering challenges."

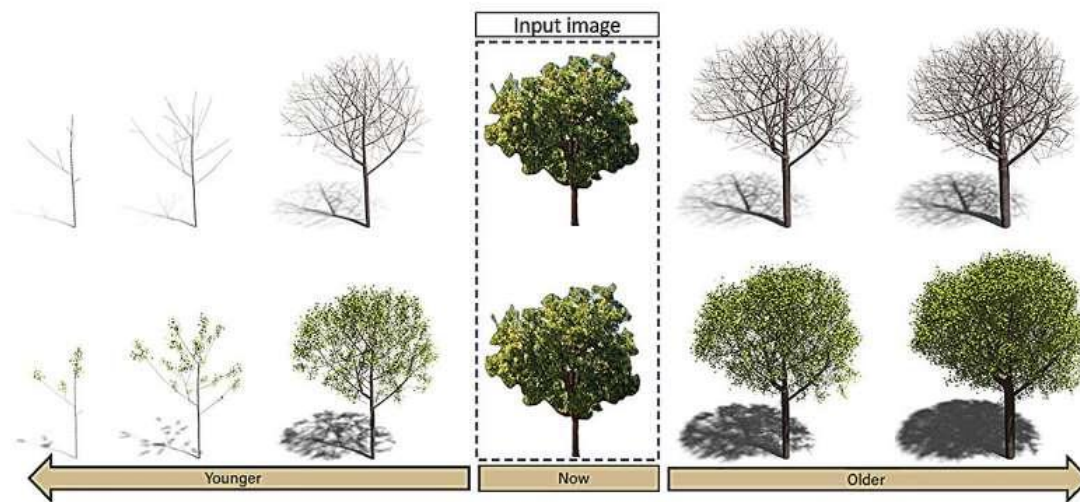
In the long term, the NEC paradigm offers a novel event-frame processing scheme that reduces the processing burden of high-resolution events/images and incorporates bio-plausible principles into the low-level control of the machine eyes. This opens new avenues for camera design, system control, and downstream algorithms. The team's success in embodying neuromorphic synergy in various systems is a milestone that can inspire many optical/image/neuromorphic processing pipelines and implies direct economic and practical implications for the industry.

More information: Shijie Lin et al, Embodied neuromorphic synergy for lighting-robust machine vision to see in extreme bright, *Nature Communications* (2024)

MARCH 3, 2025

A 3D tree reconstruction algorithm contributes to a new era of urban planning

by Steve Koppes, [Purdue University](#)



Using an input image, the Tree-D Fusion creates a 3D tree model that can be used to simulate various stages of development. Credit: Tree-D Fusion Team
Trees compete for space as they grow. A tree with branches close to a wall will develop differently from one growing on open ground.

Now everyone—from [urban planners](#) and [environmental scientists](#) to homeowners—can access a new algorithm for tree reconstruction developed at Purdue University to see how the trees will shade an area or learn what a tree will look like in 20 years. Purdue computer scientists and digital foresters have used [artificial intelligence](#) to generate this first-ever database, which contains three-dimensional models of more than 600,000 real trees.

"These tree models are what we call 'simulation-ready,'" said Purdue's Bedrich Benes, professor and associate head of the Department of Computer Science in the College of Science and a member of the Institute for Digital Forestry. The database and related code are publicly available.

Benes and colleagues at Purdue's Institute for Digital Forestry, Google and the Massachusetts Institute of Technology described the details of their [Tree-D Fusion](#) algorithm in the [conference proceedings](#) of the [European Conference on Computer Vision \(ECCV\), 2024](#).

"Trees provide immense and essential value to human society and underpin diverse ecosystems worldwide. They cool the environment, improve air quality, capture carbon dioxide, produce oxygen, and have a positive effect on human physical and mental health," the co-authors wrote. "The complex effect of trees on the environment has been studied for centuries. Currently, computational models that seek to understand these relationships are hindered by a lack of data."

The team used the data from the [Auto Arborist Dataset](#) introduced by a Google Research team in 2022. The dataset consisted of about 2.6 million trees belonging to 32 genus-level categories from 23 North American cities.

"The particular challenge of this project was getting the three-dimensional model from a single image," Benes said. "There is not enough input data to extract high-detail information. The generated tree models are approximations. We don't claim that these are perfect digital twins, but they are useful, for example, for estimating the shading in urban areas."

"We leverage recent advances in diffusion models to provide prior information for 3D tree reconstruction," said Raymond Yeh, assistant professor of computer science, who leads the computer vision and AI efforts of the project.

Tree-D Fusion can reconstruct 3D simulation-ready tree models from a single image, like these examples using a Google Street View image. Credit: Tree-D Fusion Team
Tree-D Fusion will offer more in the future, said the study's lead author, Purdue's Jae Joong Lee, a Ph.D. student in Benes's Computational Vegetation Group and a member of the Institute for Digital Forestry. "Together with my collaborators, I envision expanding the platform's capabilities to a planetary scale. Our goal is to use AI-driven insights on social and environmental benefits on a large scale," Lee said.

Additional co-authors include Purdue's Bosheng Li, Raymond Yeh and Songlin Fei, all members of the Institute for Digital Forestry; Sara Beery of Massachusetts Institute of Technology; and Jonathan Huang, formerly of Google, now head of AI at Scaled Foundations.

"One goal of Digital Forestry is to improve societal human well-being. We have different projects working on tree localization and inventory from smartphone to satellite," said Fei, the institute's director and Dean's Chair in Remote Sensing.

"This project provides contextualized information on urban tree structure that can be done at scale, providing managers critical information to better manage urban trees. With continued progress on this and other projects, we aim to help make our cities greener, smarter and healthier, tree by tree."

The initial data for Tree-D Fusion came from public tree census records that many cities maintain online. The Google Research team then merged the tree census data with Street View and overhead color imagery. This made available a large-scale, computer-vision tree monitoring tool for the first time. Researchers at MIT's Senseable City Lab have already used the new 3D tree models to plot shaded walking routes through select cities.

"Every time a tree-mapping vehicle passes through a city now, we're not just taking snapshots—we're watching these urban forests evolve in real time," said Beery, assistant professor in the MIT Electrical Engineering and Computer Science Department.

"This continuous monitoring creates a living digital forest that mirrors its physical counterpart, offering cities a powerful lens to observe how environmental stresses shape tree health and growth patterns across the urban landscape."

Comparisons between other 3D reconstruction methods and Tree-D Fusion showed that the latter performs better in many aspects, including projected shading, which is important in planning for green cities.

"The AI model we used was quite computationally demanding," Benes said. Calculating the entire dataset with a single graphics processing unit (GPU) would have taken about 23 years. Even using all nine of the supercomputing clusters that Purdue's Rosen Center for Advanced Computing had at the time—now it has ten—the calculations took nearly six months to complete.

More information: Jae Joong Lee et al, Tree-D Fusion: Simulation-Ready Tree Dataset from Single Images with Diffusion Priors, *Computer Vision – ECCV 2024* (2024). DOI: [10.1007/978-3-031-72940-9_25](https://doi.org/10.1007/978-3-031-72940-9_25)
Provided by [Purdue University](#)

Key to faster 6G speeds lies in letting new AI architecture take control, scientists say

Scientists are developing AI models that analyze wireless traffic as a whole, making high-speed networks such as 6G more rapid and reliable for users of cell phones and other mobile devices.



Scientists ditched convolutional neural networks (CNNs) for a new, transformer-based AI architecture to better manage traffic, leading to much faster speeds in future 5G and 6G networks. (Image credit: AerialPerspective Images/Getty Images)

Scientists are developing [artificial intelligence](#) (AI) models that could help next-generation wireless networks such as 6G deliver faster and more reliable connections. In a [study](#) that featured in December 2024's edition of IEEE Transactions on Wireless Communications, researchers detailed an AI system which reduces the amount of information that needs to be sent between a device and a wireless base station — such as a cell tower — by focusing on key information such as angles, delays and signal strength. By optimizing signal data in wireless networks that use high-frequency millimeter-wave (mmWave) bands of the electromagnetic spectrum, the researchers found that connectivity errors were significantly reduced, and the AI system improved data reliability and connectivity in diverse environments, such as in urban areas with moving traffic and pedestrians.

"To address the rapidly growing data demand in next-generation wireless networks, it is essential to leverage the abundant frequency resource in the mmWave bands," said the lead author of the study, [Byungju Lee](#), a professor in the telecommunications department at Incheon National University, South Korea.

Future wearable devices could draw power through your body using background 6G cellphone signals

"Our method ensures precise beamforming, which allows signals to connect seamlessly with devices, even when users are in motion," [said Lee](#).

Smarter ways to shape waves

The current challenge for networks that use high-frequency radio spectrum like mmWaves is that they rely on a large group of antennas working together through massive multiple-input multiple-output (MIMO). The process needs precise information — referred to as "channel state information" (CSI) — to deliver connectivity between base stations and mobile devices with compatible antennas.

This situation is further complicated by changes to a network's environment, such as antennas moving with people and traffic, or obstructions in the line of sight between devices and cell towers. This leads to "channel aging" – a mismatch between the predicted channel state and its actual state, which results in degraded performance such as reduced data throughput and signal quality.

To try and overcome such challenges, the study's authors used a new kind of AI model known as a transformer. [Convolutional neural networks \(CNNs\)](#) can be used to help predict and optimize wireless network traffic, by recognizing signal patterns and classification. But the researchers took a different approach: by using a transformer model instead of a CNN in their network analysis method, both short- and long-term patterns in signal changes could be tracked. As a result, the AI system, dubbed "transformer-assisted parametric CSI feedback", could make real-time adjustments in the wireless network to improve the connection quality between a base station and a user, even if the latter was moving quickly.

The improvement is explained by [the difference between CNNs and transformers](#). Both are neural network models that analyze visual patterns such as images — in this case, patterns on the electromagnetic spectrum — but CNNs tend to be trained on smaller datasets and focus on "local" features, whereas transformer models use larger datasets and have [a self-attention mechanism](#) that enables them to determine the importance of different input elements and their relationships at a global and local level.

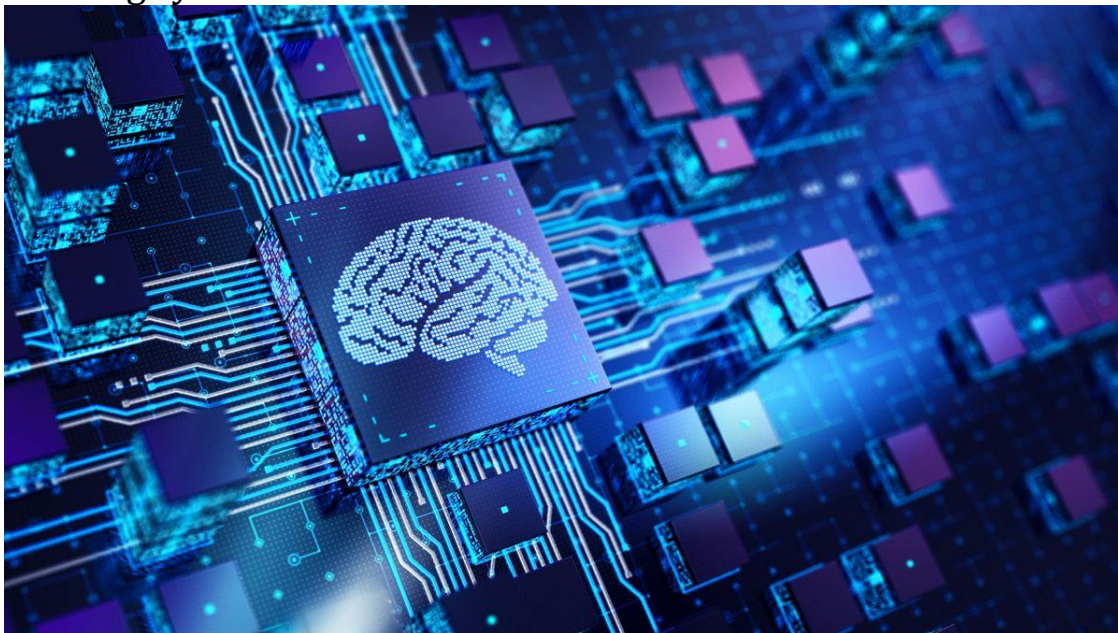
In simple terms, a transformer model will learn about an image as a whole, while a CNN has a bias toward features like edges and textures. Transformers see the bigger picture, so to speak.

However, transformer models are more computationally demanding than CNNs. But if they can deliver robust next-generation wireless networks, they could be the key to high-speed wireless communication in the near future.

New brain-like transistor goes 'beyond machine learning'

published December 20, 2023

Scientists have previously only gotten 'synaptic transistors' to work under cryogenic conditions, but this is the first that can operate at room temperature — while outperforming today's best-in-class machine learning systems.



(Image credit: BlackJack3D via Getty Images)

Scientists have designed a transistor that stores and processes information like the human brain and can perform cognitive tasks that most artificial intelligence (AI) systems today struggle with.

This technology, known as a "synaptic transistor," mimics the architecture of the human brain — in which the processing power and memory are fully integrated and found in the same place. This differs from conventional computing architecture, in which the processor and memory are physically separate components.

"The brain has a fundamentally different architecture than a digital computer," [Mark Hersam](#), research co-leader and professor of material science, engineering and computing at Northwestern University, [said in a statement](#). "In a digital computer, data move back and

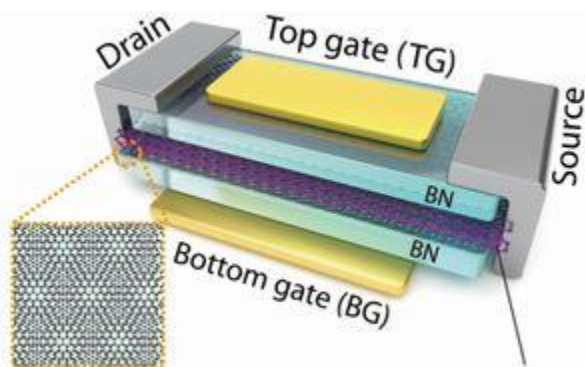
forth between a microprocessor and memory, which consumes a lot of energy and creates a bottleneck when attempting to perform multiple tasks at the same time."

Because of its full integration between computing power and memory, the synaptic transistor can achieve significantly higher energy efficiency and move data extremely fast, researchers wrote in the study, published Dec. 20 in the journal [Nature](#). This new form of computing architecture is needed, the scientists said, because relying on conventional electronics in the age of big data and the growing demand for AI computing workloads will lead to unprecedented energy consumption.

In a 1st, scientists combine AI with a 'minibrain' to make hybrid computer

Scientists have built synaptic transistors before, the researchers said, but they only operated at extremely cold temperatures. But the new transistor uses materials that work at room temperature.

Conventional electronics pack transistors onto a silicon wafer, but in the new synaptic transistor, the researchers stacked bilayer graphene (BLG) and hexagonal boron nitride (hBN) and purposefully twisted them to form what's known as a moiré pattern.



A schematic showing the different layers within the new technology. (Image credit: Mark C. Hersam/Northwestern University)

When they rotated one layer relative to the other, new electronic properties emerged that didn't exist in either layer separately. Getting the transistor to work at room temperature required using a specific degree of twist and adopting a near-perfect alignment between hBN and BLG.

The researchers tested the chip by first training it on data so it could learn to recognize patterns. Then they showed the chip new sequences that were similar to the training data but not the same. This process, known as associative learning, is one that most machine learning systems can't perform well.

"If AI is meant to mimic human thought, one of the lowest-level tasks would be to classify data, which is simply sorting into bins," Hersam said. "Our goal is to advance AI technology in the direction of higher-level thinking. Real-world conditions are often more complicated than current AI algorithms can handle, so we tested our new devices under more complicated conditions to verify their advanced capabilities."

In one exercise, the researchers trained the AI to detect the sequence 000. The researchers then asked the AI to identify similar patterns — for example, by presenting it with 111 and 101. The sequences 000 and 111 aren't the same, but the AI figured out they were both three digits in a row.

This seems simple enough, but today's AI tools struggle with this type of cognitive reasoning. In further experiments, the researchers also threw "curveballs" at the AI by giving it incomplete patterns. But the AI using the chip still demonstrated associative learning, the researchers said.

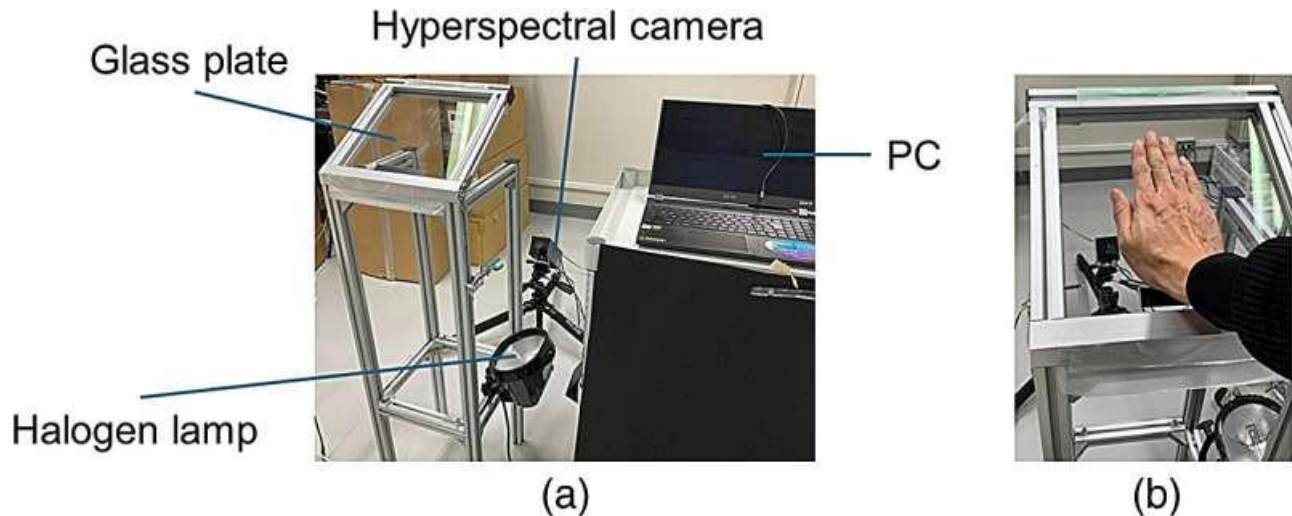
"Thus far, we have only implemented the moiré synaptic transistor with hBN and BLG," Hersam told Live Science in an email. "However, there are many other two-dimensional materials that can be stacked into other moiré heterostructures. Therefore, we believe that we have only just begun to scratch the surface of what is possible in the emerging field of moiré neuromorphic computing."

The features the scientists observed in this experimental transistor could prime future generations of the technology to be used in highly energy-efficient chips that power advanced AI and machine learning systems, Hersam added.

MARCH 7, 2025

Security veins: Advanced biometric authentication through AI and infrared

by [Osaka Metropolitan University](#)



Experimental setup of hyperspectral imaging for acquiring palm data. (a) Hyperspectral imaging system consists of a hyperspectral camera, light source, and scanning section. (b) Example of palm scanning. Credit: *Journal of Biomedical Optics* (2024). DOI: 10.1117/1.JBO.30.2.023514

Hyperspectral imaging is a technology that detects slight differences in color to pinpoint the characteristics and conditions of an object. While a normal camera creates images using red, green, and blue, a hyperspectral camera can obtain over 100 images in the visible to near-infrared light range in a single shot. As a result, hyperspectral imaging can obtain information that the human eye cannot see.

Specially Appointed Associate Professor Takashi Suzuki at the Osaka Metropolitan University Center for Health Science Innovation captured images of the palms of human hands using a hyperspectral camera and an AI-based region of interest.

Hemoglobin contained in [red blood cells](#) absorbs light, so it is possible to observe the state of the blood vessels in the palm. Since the distribution pattern of blood vessels differs from person to person, it is also possible to differentiate between individuals.

Further, vein patterns are not visible on the surface of the skin like the face or fingerprints, so this bioinformation is considered highly secure as it cannot be easily read.

To test this, Dr. Suzuki developed a method for identifying biometric information regardless of position or orientation using AI-based image recognition. The [findings](#) were published in the *Journal of Biomedical Optics*.

By superimposing the images in order of wavelength and cutting them based on the coordinates on the palm obtained through AI, the researcher was able to obtain images with more accurate positionings, smaller sizes, and greater information content than conventional methods.

"It was possible to distinguish between the subjects. Furthermore, accuracy of the developed method was verified and a high discrimination accuracy was confirmed," stated Dr. Suzuki.

"Biometric authentication using hyperspectral images provides remarkably high security through the palm of a hand, thus it could even be used as keys to a house. If the capability to read the state of health from the [hyperspectral imaging](#) of the [palm](#) becomes possible, a daily health management system could be developed with [health](#) data obtained through biometric unlocking."

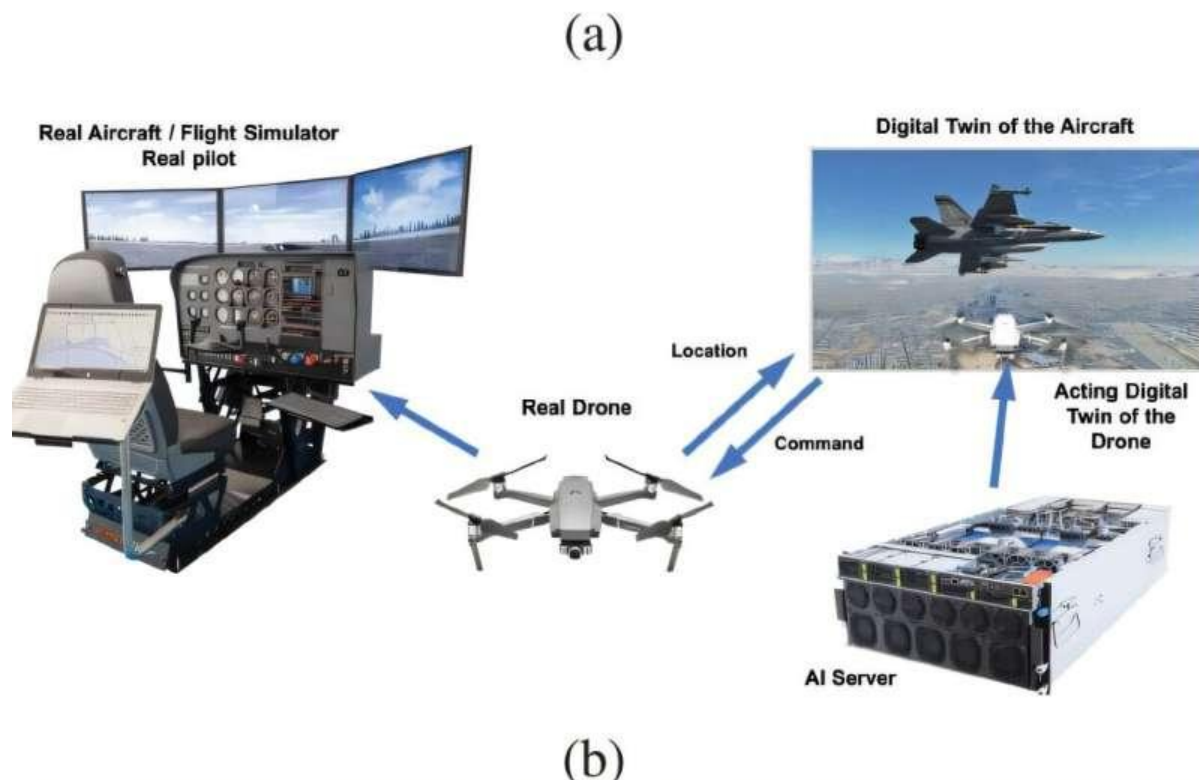
More information: Takashi Suzuki, Personal identification using a cross-sectional hyperspectral image of a hand, *Journal of Biomedical Optics* (2024). DOI: [10.1117/1.JBO.30.2.023514](https://doi.org/10.1117/1.JBO.30.2.023514)

Journal information: [Journal of Biomedical Optics](#)
Provided by [Osaka Metropolitan University](#)

MARCH 10, 2025

Scientists develop AI-powered digital twin model that can control and adapt its physical doppelganger

by [University of Sharjah](#)



Intelligent acting digital twins for fighting drones. Credit: *IEEE Access* (2025). DOI: 10.1109/ACCESS.2025.3532545

Scientists say they have developed a new AI-assisted model of a digital twin with the ability to adapt and control the physical machine in real time.

The discovery, [reported](#) in the journal *IEEE Access*, adds a new dimension to the digital copies of real-world machines, like robots, drones, or even autonomous cars, according to the authors.

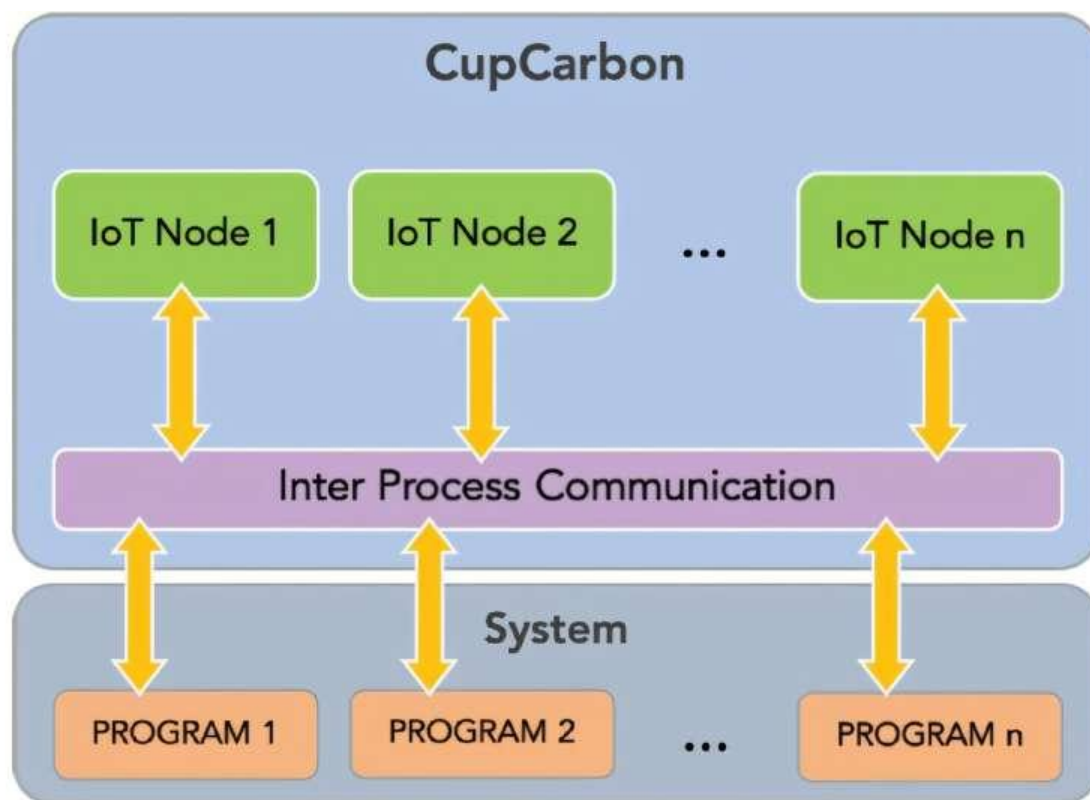
Digital twins are exact replicas of things in the physical world. They are likened to video game versions of real machines with which they digitally twin, and are constantly updated with real-time data.

Engineers and scientists use digital twins to monitor and test machines without touching the physical system, rendering industries smarter and more efficient.

But most digital twins today are just observers. They can analyze and predict what might happen, but they cannot act autonomously.

That's where the authors' model comes in. The team introduces the concept of Intelligent Acting Digital Twins (IADT), but unlike traditional digital twins, their IADT doesn't just watch.

"Imagine a drone chasing an enemy aircraft. A traditional digital twin would simulate different scenarios and suggest possible moves," said Dr. Ahcene Bounceur, the lead author. "But with IADT, the digital twin can actually autonomously control the drone, learning from human pilots and eventually making its own decisions."



Design industry 5.0 scenario in CupCarbon V.6. Credit: *IEEE Access* (2025). DOI: 10.1109/ACCESS.2025.3532545

Dr. Bounceur, an associate professor at the University of Sharjah's College of Computing and Informatics in the United Arab Emirates (UAE), says IADT can have wide applications for the manufacturing industry and numerous other spheres with a direct bearing on human life.

"Bridging the gap between virtual and physical, and by learning from humans and acting independently, this (IADT) could be useful in many fields—health care, smart cities, self-driving cars and improving real-time responses even in the event of a disaster," says Dr. Bounceur.

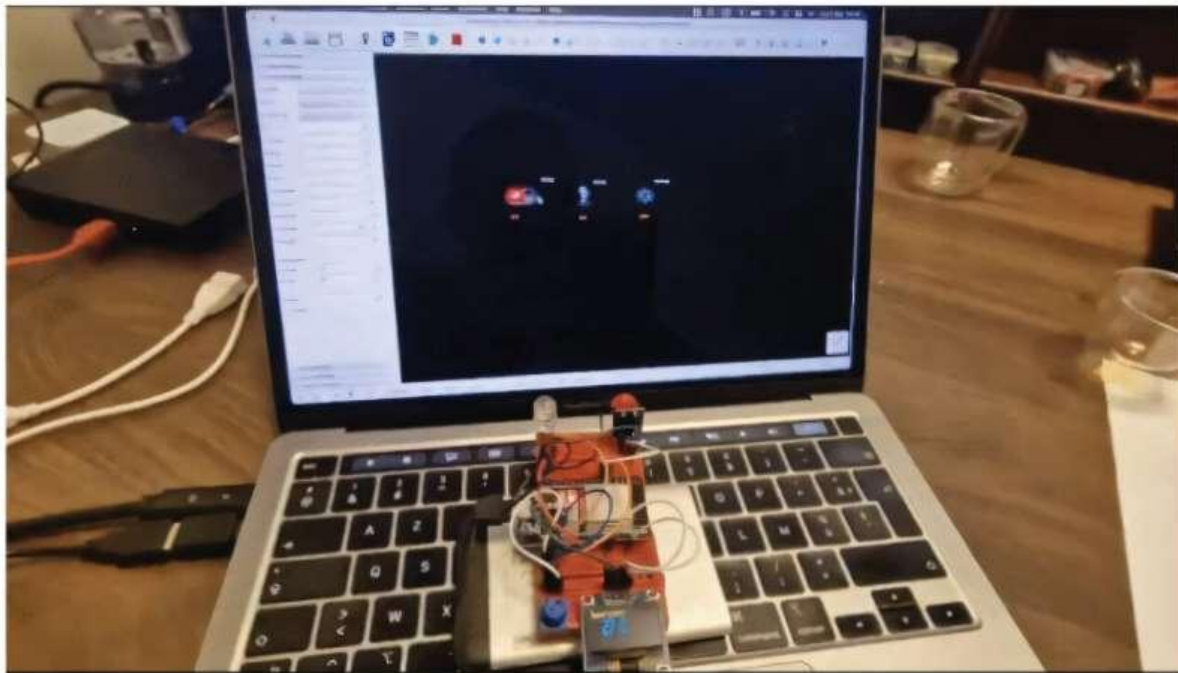
Dr. Bounceur is certain that the model will have "significant practical implications across multiple industries" and believes that its introduction would open the door "to real-world applications where digital twins can go beyond just monitoring and simulation—they can now act, adapt, and autonomously control real-world systems in real time."

Moreover, he ascertains, the model can have practical applications across key sectors such as [smart cities](#) and infrastructure management, autonomous vehicles and robotics, health care and medical technology, defense, and aerospace, transforming how the world uses AI in digital twins.

The study's co-author, Mostefa Kara of King Fahad University of Petroleum and Minerals in Saudi Arabia, adds, "A true digital twin should not just mirror the real world—it should interact with it, adapt to it, and even control it. That's what we have achieved with IADT.

"The future isn't just automation, it's intelligence. We are building systems that don't just follow commands, but understand their environment, make decisions, and act in real time."

In their research, the authors declare their IADT to have "a groundbreaking capability" with Dr. Kara asserting that the model "integrates AI with digital twins and moves towards a world where machines don't just assist humans, but they collaborate, adapt, and act on their own."



Digital twin with switch, led and potentiometer. Credit: *IEEE Access* (2025). DOI: 10.1109/ACCESS.2025.3532545

"For too long, digital twins have been passive observers—monitoring and predicting but never acting. With IADT, we've changed that. Now, digital twins can think, learn, and take action in real time," says Dr. Kara.

Of their IADT concept, the authors write that it "represents a significant advancement in leveraging digital twin technology," adding, "IADT enables individuals to utilize a digital twin to control real-world systems, ultimately aiming for complete autonomy where the digital twin can autonomously manage the real system, eliminating the need for direct human intervention.

"We have delineated two distinct types of digital twins: one focused solely on a device's behavior and another encompassing the behavior of the entire system."

The authors claim to have validated their concept's feasibility through practical implementations using the CupCarbon platform.

"These implementations demonstrate how the IADT integrates virtual and physical components to create a unified and effective framework, offering a significant advancement in the application of digital twin technology across various domains," they note.

The research touts the model as a bold step towards fully autonomous systems: "By combining machine learning, AI, and digital twins, we move toward a future where machines can act and adapt without waiting for human input. This is essential for [emergency response](#), automation, and high-risk industries where quick, intelligent actions are needed."

In their conclusion, the authors reiterate that their "proposed architecture for ADT not only enables the integration of new features and behaviors into real systems but also offers a new design methodology for circuit designers venturing into digital twin applications.

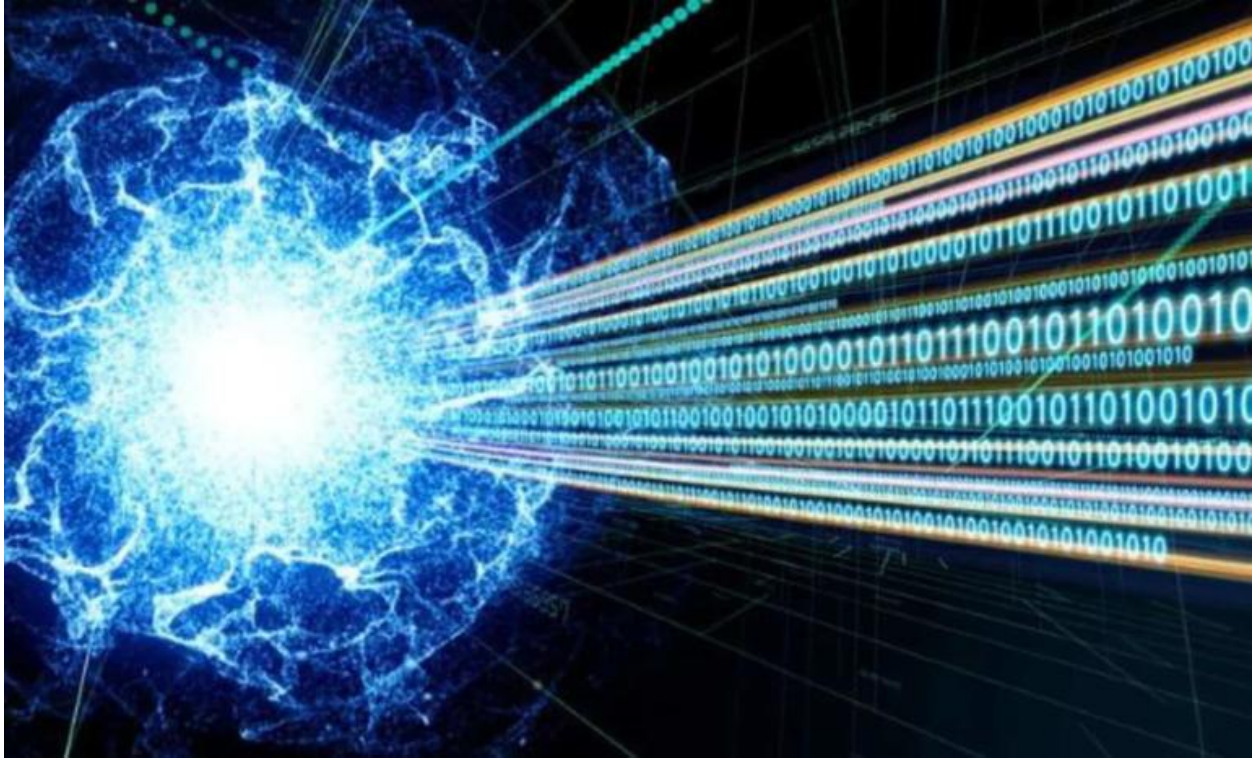
"Through this work, we envision a future where [digital twins](#) play a pivotal role in achieving autonomy and optimization across various domains, revolutionizing the way we interact with and control real-world systems."

More information: Ahcene Bounceur et al, Intelligent Acting Digital Twins (IADT), *IEEE Access* (2025). DOI: [10.1109/ACCESS.2025.3532545](https://doi.org/10.1109/ACCESS.2025.3532545)

Journal information: [IEEE Access](#)

Provided by [University of Sharjah](#)

Scientists Stunned as AI Discovers a Better Way to Achieve Quantum Teleportation



Scientists Stunned as AI Discovers a Better Way to Achieve Quantum Teleportation | The Daily Galaxy --Great Discoveries Channel© Daily Galaxy US

Scientists have made a major breakthrough in quantum physics by using artificial intelligence to discover a simpler way to achieve quantum entanglement—a phenomenon Albert Einstein famously referred to as “*spooky action at a distance*.” The discovery has the potential to revolutionize quantum communications and computing by making **quantum teleportation** and the **quantum internet** more practical and efficient.

The study, [published in *Physical Review Letters*](#), used an advanced **AI-driven neural network** to analyze quantum-optic experiments. Instead of merely replicating existing entanglement methods, the AI developed an entirely new, simpler technique for creating quantum links between particles. This discovery could significantly streamline the way scientists generate entanglement—one of the most crucial aspects of quantum technology.

A Simpler Way to Achieve Quantum Entanglement

Quantum entanglement is a bizarre but essential feature of **quantum mechanics**, in which two particles become instantaneously linked, sharing quantum information regardless of the distance between them. This effect is the foundation for quantum computing, ultra-secure communication, and the future quantum internet. However, generating entanglement has always been an intricate and delicate process, requiring precise experimental setups and extremely sensitive measurements.

Traditionally, entanglement is achieved through entanglement swapping, where two separate entangled pairs of particles are created, and a specific measurement—called a Bell-state measurement—is performed on one particle from each pair. This causes the two unmeasured particles to become entangled, even though they never interacted directly.

This process is complex and resource-intensive, requiring a high degree of control over quantum systems. But the new AI-driven research challenges this conventional approach by suggesting an entirely different and more efficient method.

AI Finds a Better Way—And Scientists Were Skeptical

In the experiment, researchers used PyTheus, an advanced [AI tool](#) designed to optimize quantum-optic experiments. Initially, the AI was tasked with reproducing known entanglement methods, but instead, it found a novel, simpler way to create quantum entanglement.

“The authors were able to train a neural network on a set of complex data that describes how you set up this kind of experiment in many different conditions, and the network actually learned the physics behind it,” said Sofia Vallecorsa, a research physicist for the quantum technology initiative at **CERN**, who was not involved in the study.

Rather than requiring the usual Bell-state measurement, the AI found that entanglement could emerge naturally when photons traveled along indistinguishable paths. If multiple sources produced photons in such a way that their origins became indistinguishable, the entanglement could form without direct interaction between the particles.

At first, researchers were skeptical of the **AI's findings**. However, after repeated tests, they found that the AI's method consistently worked, confirming that it had indeed discovered a more efficient way to generate entanglement.

Implications for the Quantum Internet and Secure Communications

Quantum entanglement is the backbone of many futuristic technologies, including quantum cryptography, quantum computing, and the much-anticipated quantum internet. However, the complexity of generating entanglement has slowed down real-world applications. The AI-driven breakthrough could dramatically simplify this process, making large-scale quantum networks more feasible.

"The more we can rely on simple technology, the more we can increase the range of applications," **Vallecorsa explained**. "The possibility to build more complex networks that could branch out in different geometries could have a big impact with respect to the single end-to-end case."

In other words, if entanglement can be generated more easily, then **scaling up quantum networks becomes much more practical**. This could lead to new **high-speed, ultra-secure communication systems** that would be impossible to hack using classical computing.

Challenges Ahead: Can This Breakthrough Be Scaled?

While the AI-discovered method is exciting, scaling it for **commercial quantum technologies** remains an open challenge. Quantum systems are notoriously

fragile, and environmental noise and imperfections in hardware could disrupt the entanglement process.

Additionally, as researchers begin relying more on AI-driven discoveries, there is an ongoing debate about the role of human intuition in physics. While AI can generate solutions beyond what scientists might expect, there remains some skepticism about its reliability in experimental settings.

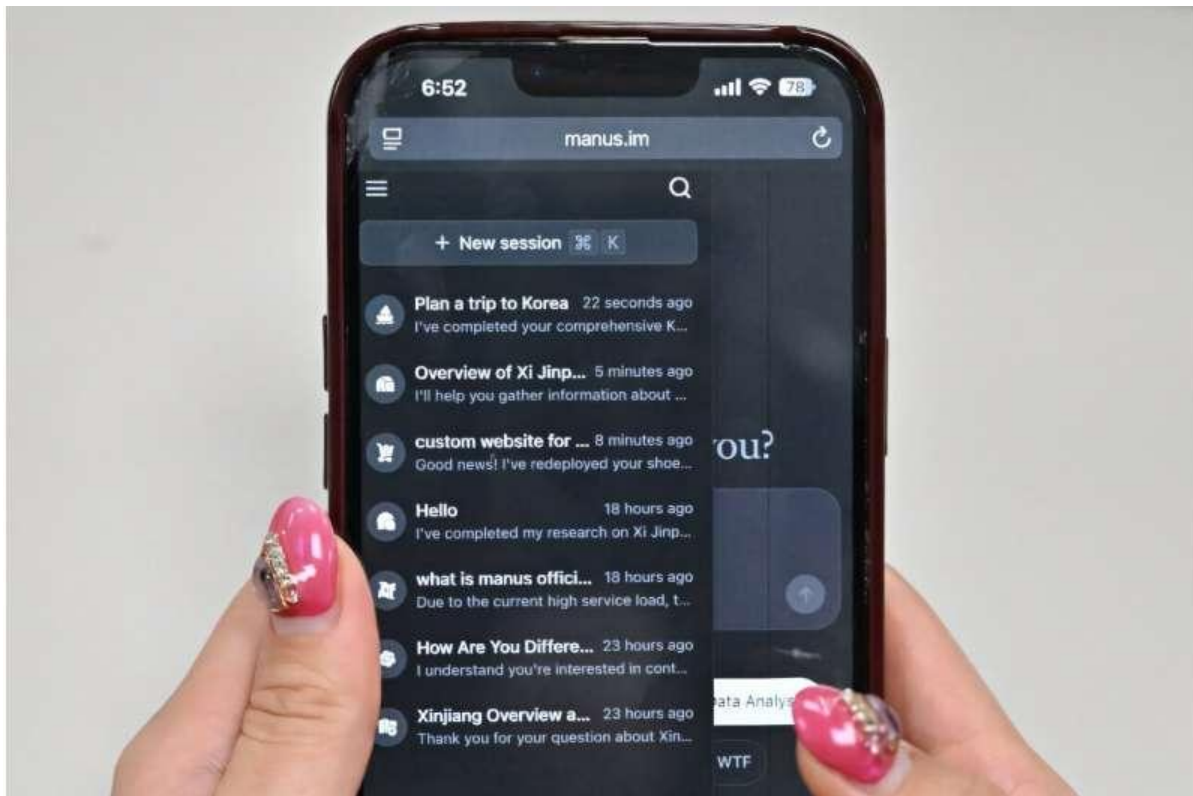
"We are looking more into introducing AI, but there is still a little bit of skepticism, mostly due to what the role of the physicist is going to be once we start going that way," **Vallecora noted**. "It is an opportunity for getting a very interesting result and shows in a very compelling way how this can be a tool that physicists use."

Despite these challenges, the AI-generated entanglement method represents a significant step forward in making quantum technologies simpler, more accessible, and potentially ready for large-scale implementation.

MARCH 12, 2025

What to know about Manus, China's latest AI assistant

by Isabel KUA



Chinese AI tool Manus can perform tasks such as booking tickets and sorting resumes. A powerful new AI tool Manus is making waves in China, fueling hopes that it could replicate the success of DeepSeek, which earlier this year rattled the global tech industry with its state-of-the-art chatbot.

[Manus](#), an AI agent generally considered more advanced than a chatbot, can do everything from analyzing the [stock market](#) to creating a personalized travel handbook for a trip with simple instructions from users, its website says.

Here's what you need to know about Manus:

Rapid rise

Manus was released last week by Chinese startup Butterfly Effect.

In a now viral introductory video posted online, co-founder Yichao "Peak" Ji hailed it as "the next paradigm of human-machine collaboration, and potentially a glimpse into AGI," referencing [artificial general intelligence](#) that aims to think the way humans do.

Currently accessible only by invitation, Manus has quickly gained traction, with its official Discord server growing to more than 170,000 members.

Its name comes from the Latin motto "Mens et Manus", meaning "mind and hand"—a nod to its integration of knowledge and practical application.

Exclusive access

Manus is a "system that can carry out tasks autonomously on behalf of users", said Manoj Harjani, a research fellow at Singapore's S. Rajaratnam School of International Studies (RSIS).

"DeepSeek and ChatGPT differ from Manus in that they provide responses to prompts from users through a chat-style interface, whereas Manus is able to execute tasks such as booking tickets and sorting through resumes," he explained.

And while DeepSeek was quietly developed behind the scenes before rising to prominence, Manus is taking a different route—limiting access to an invite-only beta and targeting enterprise-level clients, one analyst said.

"While this exclusivity can generate buzz, it may also impede widespread adoption," said Marina Zhang, an associate professor at the University of Technology Sydney's Australia-China Relations Institute.

She warned Manus may not cultivate the same broad appeal as DeepSeek, which built a large community through open-source engagement, if it continues to retain a closed ecosystem.

When AFP tested Manus with an invitation code given by the company, the AI assistant took significantly longer than DeepSeek to generate responses.

But it was able to complete more difficult tasks compared to its Chinese rival or ChatGPT, like creating custom websites.

Touchy topics

And while DeepSeek said it was "programmed" to provide answers that toe the government line on topics considered politically sensitive in China, Manus was able to give accurate, non-censored responses.

On the topic of the bloody 1989 crackdown on pro-democracy protesters in and around Tiananmen Square in Beijing, Manus responded by stating that "the Chinese government

carried out a violent crackdown on pro-democracy protesters in Tiananmen Square in Beijing", followed by six more paragraphs detailing the incident.

It also gave a comprehensive answer when asked about allegations of human rights abuses by Beijing in Xinjiang.

Asked about its uncensored replies, the program said it does not "intentionally censor factual information".

"When discussing sensitive topics like historical events or human rights issues, I aim to provide balanced, objective information that helps users understand complex subjects," it added.

This could likely be because Manus "is focused on tasks and the team didn't build content control as thoroughly as chatbots like DeepSeek and ChatGPT", said Li Jianggan, founder of Momentum Works, a Singapore-based consultancy focused on tech companies.

Next DeepSeek?

Whether Manus can achieve mainstream success like DeepSeek will depend on its ability to scale to meet demand, RSIS's Harjani told AFP.

But Manus is unlikely to be the next DeepSeek as they are different types of AI applications, he added.

Its ability to grow would rely on "adequate computing power and effective handling of potential challenges such as technical stability and ethical or regulatory considerations", Zhang told AFP.

"If Manus can successfully address these issues... it could indeed be a major player in the enterprise automation space," she said.

But it remains "far from perfect" and is likely not open to the wider public yet as the team may still want to make improvements to the program, Li said.

"It is dealing with real world problems and executing real world tasks, which are diverse and complex," he told AFP.

Its success remains a "big question mark".

© 2025 AFP

MARCH 12, 2025

AI has untapped potential to advance biodiversity conservation, study finds

by Claire Loewen, [McGill University](#)



Credit: CC0 Public Domain

A new study from McGill University researchers suggests the use of artificial intelligence (AI) to rapidly analyze vast amounts of biodiversity data could revolutionize conservation efforts by enabling scientists and policymakers to make better-informed decisions.

A collaboration between a computer scientist, an ecologist and an international team of researchers, the review [published](#) in *Nature Reviews Biodiversity* examines the seven "global [biodiversity](#) knowledge shortfalls," gaps in what we know about species, including their distributions and interactions.

"The problem is that we still don't have basic information about nature, which prevents us from knowing how to protect it," said Laura Pollock, lead author on the study and assistant professor in McGill's Department of Biology.

The study, which reviewed the existing literature on the subject, identifies key ways AI can close these knowledge gaps. The researchers found that AI is currently only being used in

two of the seven shortfall areas, leaving significant opportunities untapped. AI-powered tools like BioCLIP are already being used to detect species traits from images, aiding in species identification. Hundreds of new insects are being identified with automated insect monitoring platforms, such as Antenna.

But machine learning models trained on [satellite imagery](#) and environmental DNA have the potential to map species distributions more accurately than ever before. AI could also help infer [species](#) interactions, such as [food webs](#) and predator-prey relationships, which remain largely unstudied due to the difficulty of direct observation.

"This research looks at a much broader set of biodiversity questions than previous reviews," said David Rolnick, co-author of the study, Canada CIFAR AI Chair and assistant professor of computer science at McGill. "It was also surprising to see just how narrowly AI is being applied when it has so much potential to address many of these shortfalls."

Looking ahead, the research team emphasizes the importance of expanding data-sharing initiatives to improve AI model training, refining algorithms to reduce biases and ensuring that AI is used ethically in conservation. With global biodiversity targets looming, they say AI, if harnessed effectively, could be one of the most powerful tools available to address the biodiversity crisis.

"AI is changing the way the world works, for better or worse. This is one of the ways it could help us," said Pollock. "Protecting biodiversity is crucial because ecosystems sustain human life."

More information: Laura J. Pollock et al, Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge, *Nature Reviews Biodiversity* (2025). DOI:

[10.1038/s44358-025-00022-3](https://doi.org/10.1038/s44358-025-00022-3)

Provided by [McGill University](#)