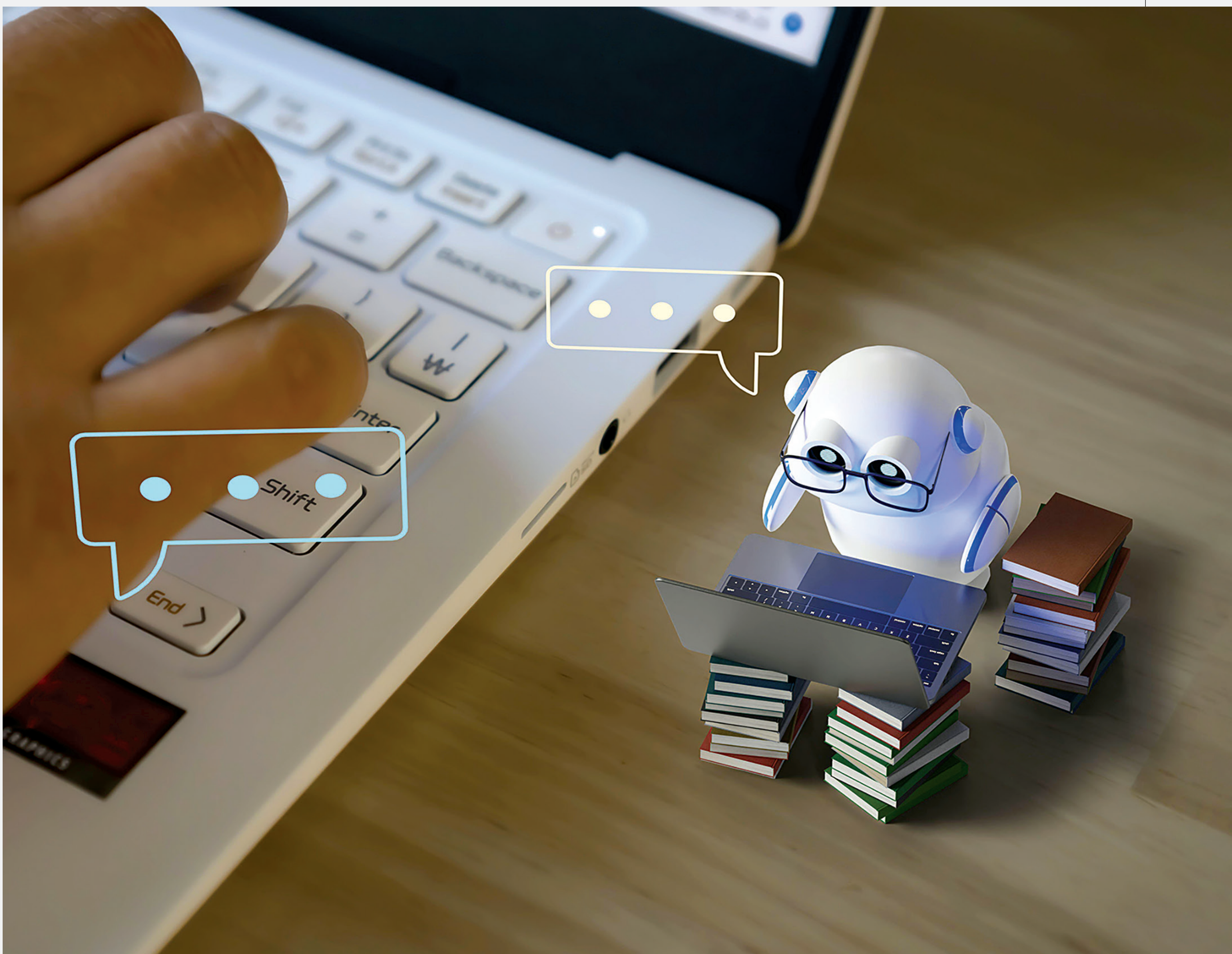
News



ALAMY

ARTIFICIAL INTELLIGENCE

Professors Rethink How They Teach Coding › Students embrace AI copilots, teachers shift to problem solving

BY RINA DIANE CABALLAR

It's clear that generative AI is transforming the software-development industry. AI-powered coding tools are assisting programmers in their workflows, while the number of jobs in AI continues to increase. But the shift is also evident in academia—one of the major avenues through which the next generation of software engineers is learning how to code.

Computer science students are embracing the technology, using generative AI to help them understand complex concepts, summarize complicated research papers, brainstorm ways to solve a problem, come up with new research directions, and, of course, learn how to code.

"Students are early adopters and have been actively testing these tools," says Johnny Chang, a teaching assistant at Stanford University pursuing a master's degree in computer science. He founded the AI x Education conference in 2023, a virtual gathering of students and educators to discuss the impact of AI on education.

Educators are also experimenting with generative AI. But they're grappling with techniques to adopt the technology while still ensuring students learn the foundations of computer science.

"It's a difficult balancing act," says Ooi Wei Tsang, an associate professor in the School of Computing at the National University of Singapore. "Given that large language models are evolving rapidly, we are still learning how to do this."

The fundamentals and skills themselves are evolving. Most introductory computer science courses focus on code syntax and getting programs to run, and while knowing how to read and write code is still essential, testing and debugging—which aren't commonly part of the syllabus—now need to be taught more explicitly.

"We're seeing a little upping of that skill, where students are getting code snippets from generative AI that they need to test for correctness," says Jeanna Matthews, a professor of computer science at Clarkson University in Potsdam, N.Y.

Another vital expertise is problem decomposition. "This is a skill to know early on because you need to break a large problem into smaller pieces that an LLM can solve," says Leo Porter, an associate teaching professor of computer science at the University of California, San Diego. "It's hard to find where in the curriculum that's taught—maybe in an algorithms or software engineering class, but those are advanced classes. Now, it becomes a priority in introductory classes."

As a result, educators are modifying their teaching strategies. "I used to have this singular focus on students writing code that they submit, and then I run test cases on the code to determine what their

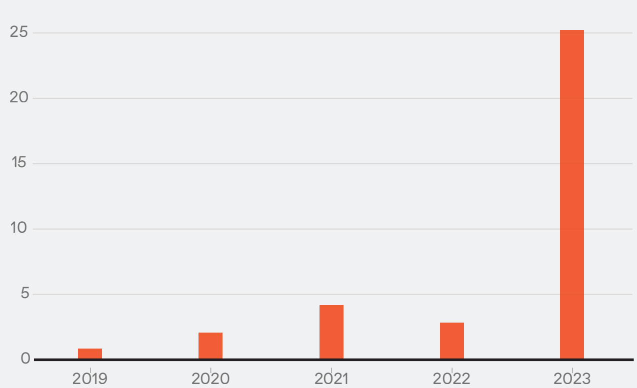
The State of AI in 2024

Each year, the AI Index lands on virtual desks with a louder virtual thud. This year, its 502 pages are a testament to the fact that we're in the midst of an AI boom. The report, published in April by the Stanford Institute for Human-Centered Artificial Intelligence (HAI), covers trends in R&D, technical performance, responsible AI, the economy, education, policy and governance, diversity, and public opinion. Here we've compiled four charts that speak to the state of AI in 2024. —**ELIZA STRICKLAND**

For our full selection of 15 charts, see https://spectrum.ieee.org/ai-index-2024

PRIVATE INVESTMENT IN GENERATIVE AI, 2019–23

TOTAL INVESTMENT, US $, BILLIONS

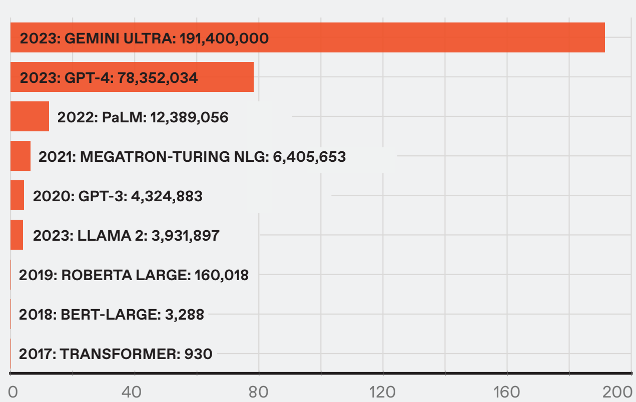


SOURCE: QUID, 2023 | CHART: 2024 AI INDEX REPORT

GENERATIVE AI INVESTMENT SKYROCKETS: While corporate investment was down overall last year, investment in generative AI went through the roof. Nestor Maslej, editor in chief of this year's report, says the boom is indicative of a broader trend as the world grappled with the new capabilities and risks of generative AI systems like ChatGPT and DALL-E 2. "Whether it's in policy, in public opinion, or in industry," says Maslej, people have responded to generative AI "with a lot more investment."

ESTIMATED TRAINING COST OF SELECT AI MODELS, 2017–23

TRAINING COST, US $, MILLIONS



SOURCE: EPOCH, 2023 | CHART: 2024 AI INDEX REPORT

FOUNDATION MODELS HAVE GOTTEN SUPEREXPENSIVE: Foundation models—big generative AI models that allow for a variety of language or art tasks—are predominately coming from industry, because training one takes very deep pockets. AI companies rarely reveal the expenses involved in model training, so the AI Index estimated costs by analyzing information gleaned from publications, press releases, and technical reports. Particularly noteworthy: Google's 2017 Transformer model, which introduced the architecture that underpins almost all of today's large language models, was trained for only US $930.

grade is," says Daniel Zingaro, an associate professor of computer science at the University of Toronto Mississauga. "This is such a narrow view of what it means to be a software engineer, and I just felt that with generative AI, I've managed to overcome that restrictive view."

Zingaro, who coauthored a book on AI-assisted Python programming with Porter, now has his students work in groups and submit a video explaining how their code works. Through these walk-throughs, he gets a sense of how students use AI to generate code, what they struggle with, and how they approach design, testing, and teamwork.

"It's an opportunity for me to assess their learning process of the whole software development [life cycle]—not just code," Zingaro says. "And I feel like my courses have opened up more, and they're much broader than they used to be. I can make students work on larger and more advanced projects."

Ooi echoes that sentiment, noting that generative AI tools "will free up time for us to teach higher-level thinking—for example, how to design software, what is the right problem to solve, and what are the solutions. Students can spend more time on optimization, ethical issues, and the user-friendliness of a system rather than focusing on the syntax of the code."

But educators are also cautious given the tendency of LLMs to hallucinate. "We need to be teaching students to be skeptical of the results and take ownership of verifying and validating them," says Matthews.

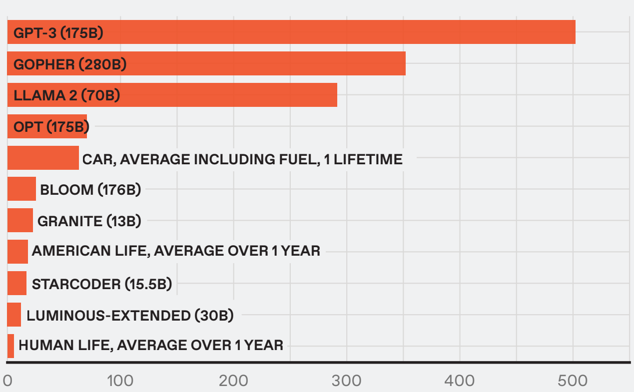
Matthews adds that generative AI "can short-circuit the learning process of students relying on it too much." Chang agrees that this overreliance can be a pitfall and advises his fellow students to explore possible solutions to problems by themselves so they don't lose out on that critical-thinking or effective learning process. "We should be making AI a copilot—not the autopilot—for learning," he says.

Other drawbacks include copyright issues and bias. "I teach my students about the ethical constraints—that this is a model built off other people's code and we'd recognize the ownership of that," Porter says. "We also have to recognize that models are going to represent the bias that's already in society."

Adapting to the rise of generative AI involves students and educators working together and learning from one another. For her colleagues, Matthews's advice is to "try to foster an environment where you encourage students to tell you when and how they're using these tools. Ultimately, we are preparing our students for the real world, and the real world is shifting, so sticking with what you've always done may not be the recipe that best serves students in this transition."

Porter is optimistic that the changes they're applying now will serve students well in the future. "There's this long history of a gap between what we teach in academia and what's actually needed as skills when students arrive in the industry," he says. "There's hope on my part that we might help close the gap if we embrace LLMs." ■

CO_2-EQUIVALENT EMISSIONS, TONNES BY SELECT MACHINE LEARNING MODELS AND REAL-LIFE EXAMPLES, 2020–23

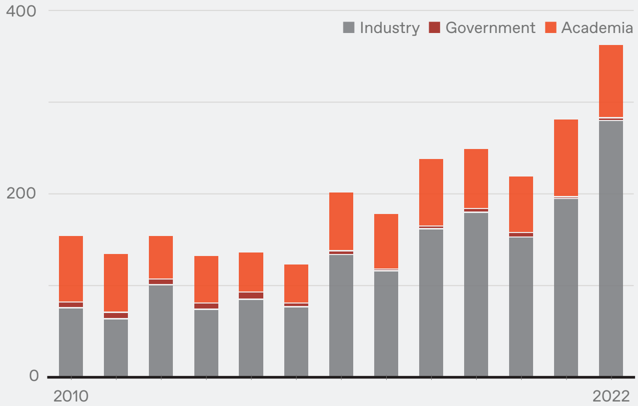


SOURCE: AI INDEX, 2024; LUCCIONI ET AL., 2022; STRUBELL ET AL., 2019 | CHART: 2024 AI INDEX REPORT

LARGE LANGUAGE MODELS HAVE A HEFTY CARBON FOOTPRINT: The AI Index team estimated the carbon footprint of training certain large language models, including such factors as model size (measured in billions of parameters), data-center energy efficiency, and the carbon intensity of energy grids. This chart does not include emissions related to models' "inference"—when they do the work they're trained for. While "per-query" emissions may be low, the report notes, the total impact can surpass training emissions if models are queried millions of times daily.

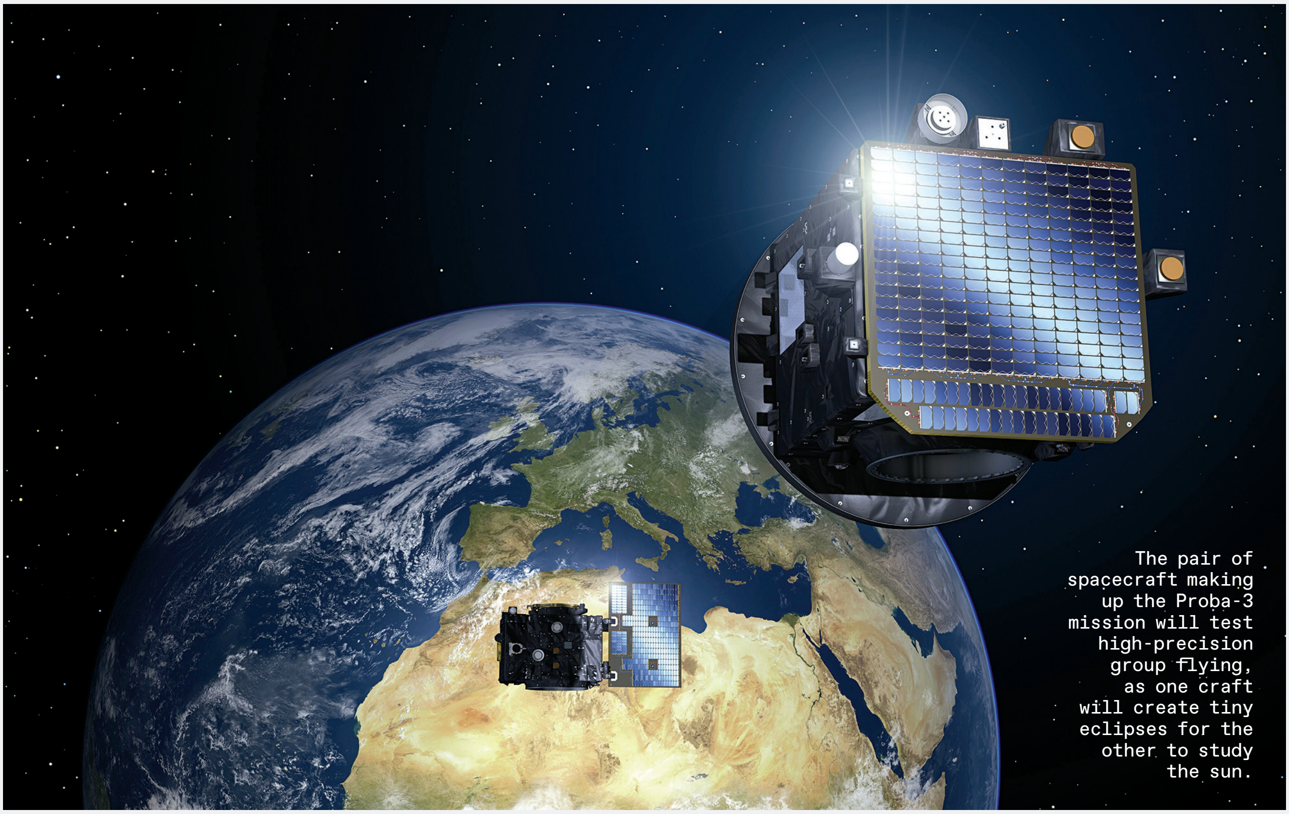
EMPLOYMENT OF NEW AI PH.D.s IN THE UNITED STATES AND CANADA BY SECTOR, 2010–22

NUMBER OF NEW AI PH.D. GRADUATES



SOURCE: CRA TAULBEE SURVEY, 2023 | CHART: 2024 AI INDEX REPORT

NEW PH.D.S FLOCK TO INDUSTRY: That more graduates of AI programs are drawn to industry is hardly a surprise, given the amount of private investment for generative AI, and the fact that industry dominates in the production of AI foundation models. In 2022 (the most recent year for which the AI Index has data), 70 percent of new AI Ph.D.s in North America took jobs in industry. It's a continuation of a trend over the last decade, and it indicates an intensifying brain drain from universities.



The pair of spacecraft making up the Proba-3 mission will test high-precision group flying, as one craft will create tiny eclipses for the other to study the sun.

P. CARRIL/ESA

AEROSPACE

Satellites Team Up to Make Mini-Eclipses › Proba-3 will demonstrate formation flying

BY ANDREW JONES

The European Space Agency will launch a mission in the coming months to create artificial solar eclipses—using two spacecraft that will position themselves in a precise formation while in orbit.

ESA's Proba-3 (PRoject for On-Board Autonomy) includes a pair of spacecraft: the 300-kilogram Coronagraph spacecraft and the 250-kg Occulter. The pair are slated to launch on a Polar Satellite Launch Vehicle rocket developed by the Indian Space Research Organization. After entering a highly elliptical orbit—600 kilometers at its closest point and 60,350 km at its farthest—the two spacecraft will be positioned roughly 144 meters apart. The aim, ESA says, is to use the Occulter to block out the sun from the Coronagraph's point of view.

Achieving this formation will allow the Coronagraph to study our star's highly ionized, extremely hot atmosphere. More broadly, the two craft will demonstrate precision flying as a precursor for more ambitious formation-flying endeavors in the future.

Solar eclipses, aside from being moments of stark beauty, are fleeting opportunities to conduct science. Helium was first detected in the sun's chromosphere by a French astronomer during a 1868 total solar eclipse in India. Fifty years later, British astronomers Frank Dyson and Arthur Eddington performed measurements of the apparent shift of stars during a 1919 solar eclipse as an early test of Einstein's general theory of relativity.

ESA will use Proba-3 to study solar astrophysics without any interference from Earth's atmosphere. The mission's coronagraph (aboard the craft of the same name) will help to discern why the solar corona is significantly hotter than the sun itself, which could improve solar-weather predictions.

However, Proba-3's precision-formation flying could be the mission's most significant contribution toward unlocking future aerospace breakthroughs. During each orbit, as the two craft approach their farthest point from Earth, they will align to create a roughly 6-hour mini-eclipse from Coronagraph's point of view. The two spacecraft will use radio links to communicate with one another and star trackers to determine their attitudes.

Additional coordination occurs as the spacecraft approach the closest point to Earth in their orbit. Global navigation satellite system (GNSS) receivers will pick up GPS signals, which, when combined with a dedicated relative navigation algorithm, will help determine the relative position of the two craft. Optical sensors on the Occulter will detect pulsing LEDs on the Coronagraph to provide data for finer measurements.

After all that, one final bit of technological wizardry will allow for millimeter-scale precision between the two craft. The Occulter will ping a laser at a corner cube retroreflector mounted on the Coronagraph spacecraft, which will bounce the beam back. Using the data gathered, Coronagraph can, the ESA says, control and maintain millimeter-level accuracy using 10-millinewton-scale cold gas thrusters aboard the craft.

"Guidance, navigation, and control has undergone a lot of development, and this is what we all want to demonstrate," Damien Galano, Proba-3's project manager, said during a press briefing on 3 April.

"We have an actual application, which is observation of the corona. So, by achieving really good observations of the corona, we would definitely demonstrate that all this equipment is working and that the technology is delivering actual science data," Galano said, adding that Proba-3 formation-flying control algorithms and metrology systems could be applied to future missions.

For example, precisely controlled Occulter-like spacecraft could be used with space telescopes to block light from a distant star in order to directly detect any potential orbiting exoplanets. Constellations of spacecraft could, through interferometry, create large-scale observatories by collectively achieving large apertures and long focal lengths that would be impossible with large solo satellites.

Other applications include Earth observation, space-based gravitational-wave detection, and a range of missions in which two or more spacecraft need to interact, such as rendezvous, docking, and in-orbit servicing missions.

Before such complex projects can be undertaken, Proba-3 needs to prove it can nail the basics. "That's clearly the operational challenge," said Dietmar Pilz, ESA director of technology, engineering, and quality. "To see how far we can get the formation flying, what are the distances that we can achieve. This needs a lot of operational expertise and software from all the partners in this project." ■

JOURNAL WATCH

Sea Turtles Inspire Heart-Monitor Design

RESEARCHERS ARE KEEN to develop alternative, lower-cost solutions to monitor heart health that people can use at home, in lieu of doctors using stethoscopes or echocardiograms. A team in China has now created a heart-monitoring system inspired by the auditory systems of sea turtles.

Junbin Zang, a lecturer at the North University of China, and his colleagues were intrigued when they learned about the inner workings of the sea turtle's ears, which are able to detect low-frequency signals, especially in the 300- to 400-hertz range.

"Heart sounds are also low-frequency signals, so the low-frequency characteristics of the sea turtle's ear have provided us with great inspiration," explains Zang.

At first glance, sea turtles don't seem to have ears—they lack external ear canals. Instead their auditory system lies under a layer of skin and fat through which the turtle picks up vibrations. As with humans, a small bone vibrates as sounds hit it and converts those vibrations to electrical signals that are sent to the brain for processing.

But sea turtles also have a unique slender, T-shaped conduit that encapsulates their ear bones, restricting the movement of the similarly T-shaped ear bones so that they vibrate only in a perpendicular manner. This design results in a high sensitivity.

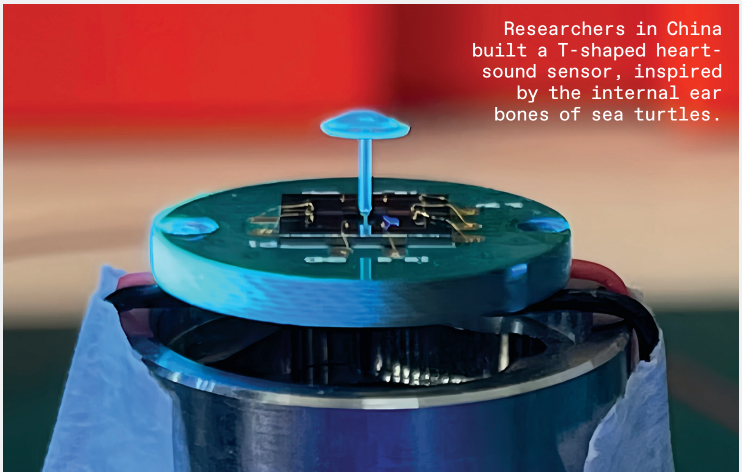
Zang and his colleagues created a similarly T-shaped heart-sound sensor using a tiny microelectromechanical systems (MEMS) cantilever beam sensor. As sound hits the sensor, the vibrations cause deformations in its beam, and the fluctuations in the voltage resistance are then translated into electrical signals.

The researchers first tested the sensor's ability to detect sound in lab tests, and then tested its ability to monitor heartbeats in two human volunteers in their early 20s. The results, described in a study published 1 April in *IEEE Sensors Journal*, show that the sensor can effectively detect the two phases of a heartbeat.

"The sensor exhibits excellent vibration characteristics," Zang says, noting that it has a higher vibration sensitivity compared with other accelerometers on the market.

However, the sensor currently picks up a significant amount of background noise, which Zang says his team plans to address in future work. Ultimately, they are interested in integrating this novel bioinspired sensor into devices they have previously created—including portable handheld and wearable versions and a relatively larger version for use in hospitals—for the simultaneous detection of electrocardiogram and phonocardiogram signals.

—Michelle Hampson



Researchers in China built a T-shaped heart-sound sensor, inspired by the internal ear bones of sea turtles.

NORTH UNIVERSITY OF CHINA

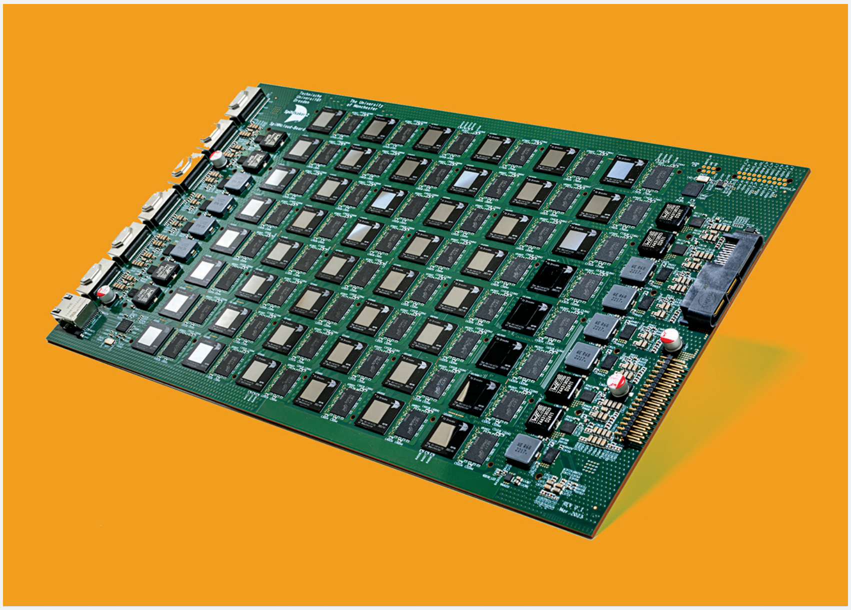
COMPUTING

Brain-Emulating Computer Reaches the Market › AI accelerators meet neuromorphic models

BY DINA GENKINA

In May, at the ISC High Performance conference in Hamburg, Germany, SpiNNcloud Systems announced that its neuromorphic (or brain-emulating) supercomputer, the SpiNNcloud Platform, was available for sale. The machine, which combines traditional AI accelerators with neuromorphic computing capabilities, is the first commercially available neuromorphic computer. The largest version of the SpiNNcloud Platform can simulate 10 billion neurons, about one-tenth the number in the human brain—a significant improvement over the previous record holder, Hala Point, which can emulate 1.15 billion neurons.

This board features 48 SpiNNaker2 chips working together to enable both neuromorphic computing and traditional AI processing.



SPINNCLOUD SYSTEMS

SpiNNcloud Systems, a startup based in Dresden, Germany, was founded in 2021 as a spin-off of the Dresden University of Technology. Its original chip, the SpiNNaker1, was designed by Steve Furber, the principal designer of the Arm microprocessor—the technology that now powers most cellphones. The SpiNNaker1 chip is already in use by 60 research groups in 23 countries, SpiNNcloud Systems says.

Neuromorphic computers promise vastly lower energy consumption for computations and better performance on certain tasks, such as training recurrent neural networks and processing sensor data in real time. "The human brain is the most advanced supercomputer in the universe, and it consumes only 20 watts to achieve things that artificial intelligence systems today only dream of," says Hector Gonzalez, cofounder and co-CEO of SpiNNcloud Systems. "We're basically trying to bridge the gap between brain inspiration and artificial systems."

Aside from the sheer number of neurons that the SpiNNcloud Platform can simulate, a distinguishing feature of the system is its flexibility. Most neuromorphic computers emulate the brain's spiking nature: Neurons fire off electrical spikes to communicate with the neurons around them. The actual mechanism of these spikes in the brain is quite complex, and neuromorphic hardware often implements a specific simplified model. However, the SpiNNcloud Platform can implement a broad range of brain-spiking models, as specific models are not hardwired into its architecture.

Instead of looking at how each neuron and synapse operates in the brain and trying to emulate that from the bottom up, Gonzalez says, SpiNNcloud's approach involves implementing key performance features of the brain. "It's more about taking practical inspiration from the brain, following particularly fascinating aspects such as how the brain is energy proportional and how it is simply highly parallel."

To build hardware that is energy proportional (meaning each portion draws power only when it's actively in use) and highly parallel, the company started with the SpiNNaker2 chips that compose the computer. Each SpiNNaker2 chip hosts 152 processing units, with each processing unit having an Arm-based microcontroller. Unlike its SpiNNaker1 predecessor, the SpiNNaker2 also comes equipped with accelerators for use on neuromorphic models and traditional neural networks.

The processing units can stay switched off unless a spike from a connected unit or an external sensor event triggers them to turn on. This enables energy-proportional operation because each unit draws power only as needed. There is no central clock to coordinate movements; instead, messages are routed between units and across chips when needed without having to wait their turn in a clock cycle. Multiple messages can be sent through different parts of the chip at the same time, enabling parallelism. Each chip is connected to six other chips, and the whole system is connected in the shape of a torus to ensure all connecting wires are equally short.

The largest commercially offered system is not only capable of emulating 10 billion neurons but also performing 0.3 billion billion operations per second of more traditional AI tasks, putting it on the scale of the top 10 largest supercomputers today.

Among the first customers of the SpiNNaker2 system is a team at Sandia National Laboratories in New Mexico, which plans to use it for further research on how neuromorphic systems can outperform traditional computer architectures and how they can perform fundamentally different computational tasks, such as implementing various models of the human brain.

For example, Fred Rothganger, senior member of technical staff at Sandia, says the flexibility of the SpiNNaker2 systems will allow Sandia researchers to study different types of artificial neural networks that may have relevance to the human brain. "[Older neuromorphic systems], of course, can run on a general-purpose computer. But those general-purpose computers are not necessarily designed to efficiently handle the kind of communication patterns that go on inside a spiking neural network. With [the SpiNNaker2 system] we get the ideal combination of greater programmability plus efficient communication." ■

SCIENCE PHOTO LIBRARY/GETTY IMAGES

BIOMEDICAL

Compact MRI Ditches Superconducting Magnets ›

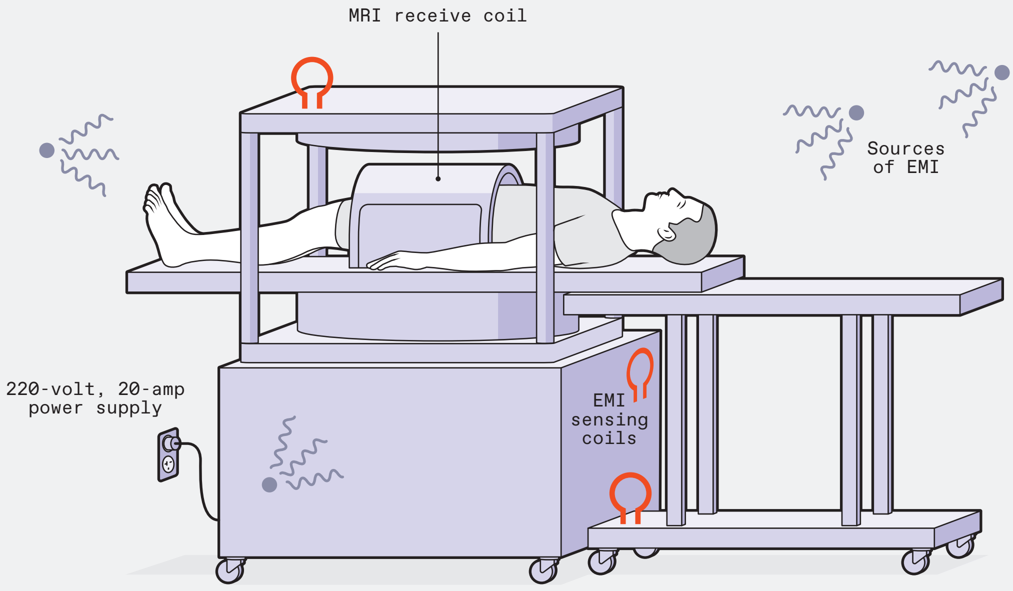
The whole-body scanner can plug into a standard wall outlet

BY CHARLES Q. CHOI

Magnetic resonance imaging (MRI) has revolutionized health care by providing radiation-free, noninvasive 3D medical images. However, most MRI scanners are room-size, complex, and power hungry—consuming 25 kilowatts or more to produce magnetic fields up to 1.5 teslas strong. As a result, scanners are typically limited to specialized departments in hospitals.

A University of Hong Kong team has unveiled a low-power, highly simplified, full-body MRI scanner. With the help of artificial intelligence, the new scanner requires only a compact 0.05-T magnet and can run off a standard wall power outlet, requiring only 1,800 watts during operation. The researchers say their new machine can produce clear, detailed images on a par with those from high-power MRI scanners currently used in clinics.

"It's the beginning of a multidisciplinary endeavor to advance an entirely new class of simple, patient-centric, and computing-powered point-of-care diagnostic imaging device," says Ed Wu, a professor and chair of biomedical engi-



This simplified MRI scanner uses weak 0.05-tesla magnets. As a result, it doesn't require expensive, bulky shielding for either its own magnets or for external sources of electromagnetic interference (EMI). Coils around the scanner detect EMI, and deep-learning models filter the interference out so that the weak magnets can still generate sharp images. The entire scanner can be powered by a standard wall outlet.

neering at the University of Hong Kong. Wu and the rest of the research team published their work on 10 May in *Science*.

More than 150 million MRI scans are conducted worldwide annually, according to the Organization for Economic Cooperation and Development. However, despite five decades of development, clinical MRI procedures remain out of reach for more than two-thirds of the world's population, especially in low- and middle-income countries. For instance, the United States has about 40 scanners per million people; as of 2021, Tanzania had fewer than one scanner per 10 million people.

These disparities largely stem from the high costs and specialized settings required for standard MRI scanners. They use powerful superconducting magnets that require a lot of space, power, and specialized infrastructure. They also need rooms shielded from radio interference, which restricts their mobility and hampers their availability in nonshielded environments.

Scientists have already been exploring low-cost MRI scanners that operate at ultralow-field (ULF) magnetic-field strengths (less than 0.1 T). Current ULF MRI scanners often rely on AI to help reconstruct images from what signals they gather using relatively weak magnetic fields. Until now, these devices were limited to solely imaging the brain, extremities, or single organs, notes Udunna Anazodo, an assistant professor of neurology and neurosurgery at McGill University in Montreal who did not take part in the work, in a review of the new study.

The Hong Kong team's whole-body ULF MRI scanner places a patient between two permanent neodymium ferrite boron magnet plates. Although these permanent magnets are far weaker than superconductive magnets, they are low-cost, readily available, and don't need to be cooled to frigid temperatures using expensive liquid helium.

The new machine consists of two units, each roughly the size of a hospital gurney. One unit houses the MRI device, while the other supports the patient's body as it slides into the scanner.

To account for radio interference from both the outside environment and the ULF MRI's own electronics, the scientists deployed 10 small sensor coils around the scanner and inside the electronics cabinet to help the machine detect potentially disruptive radio signals. The researchers also employed deep-learning AI methods to help reconstruct images even in the presence of strong noise. They say this eliminates the need for shielding against radio waves, making the new device far more portable than a conventional MRI.

In tests on 30 healthy volunteers, the device captured detailed images of individuals' brains, spines, abdomens, hearts, lungs, and extremities. Scanning each of these targets took 8 minutes or less for image resolutions of roughly 2-by-2-by-8 cubic millimeters. In Anazodo's review, she notes that the new machine produced image qualities comparable to those of conventional MRI scanners.

The researchers used standard off-the-shelf electronics. All in all, they estimate hardware costs at about US $22,000. (Entry-level MRI scanners can start at $225,000, and advanced machines can cost $500,000 or more.)

The prototype scanner's magnet assembly is heavy, weighing about 1,300 kilograms. (This is lightweight, however, compared with a typical clinical MRI scanner, which can weigh up to 17 tonnes.) The scientists note that optimizing the hardware could reduce the magnet assembly's weight to about 600 kg, which would make the entire scanner mobile.

The researchers say their new device is not meant to replace conventional high-magnetic-field MRI. A 2023 study notes that next-generation MRI scanners using powerful 7-T magnets could yield a resolution of just 0.35 millimeters. ULF MRI could instead complement existing MRI by going to places that can't host standard MRI devices, such as intensive-care units and community clinics.

In an email to *IEEE Spectrum*, Anazodo says this new Hong Kong work is just one of a number of exciting ULF MRI scanners under development. She also points to researchers at the University of Saskatchewan, in Canada, who are developing a device that is potentially even lighter, cheaper, and more portable than the Hong Kong machine, which they are researching for use in whole-body imaging on the International Space Station. ■

Illustration by Chris Philpot