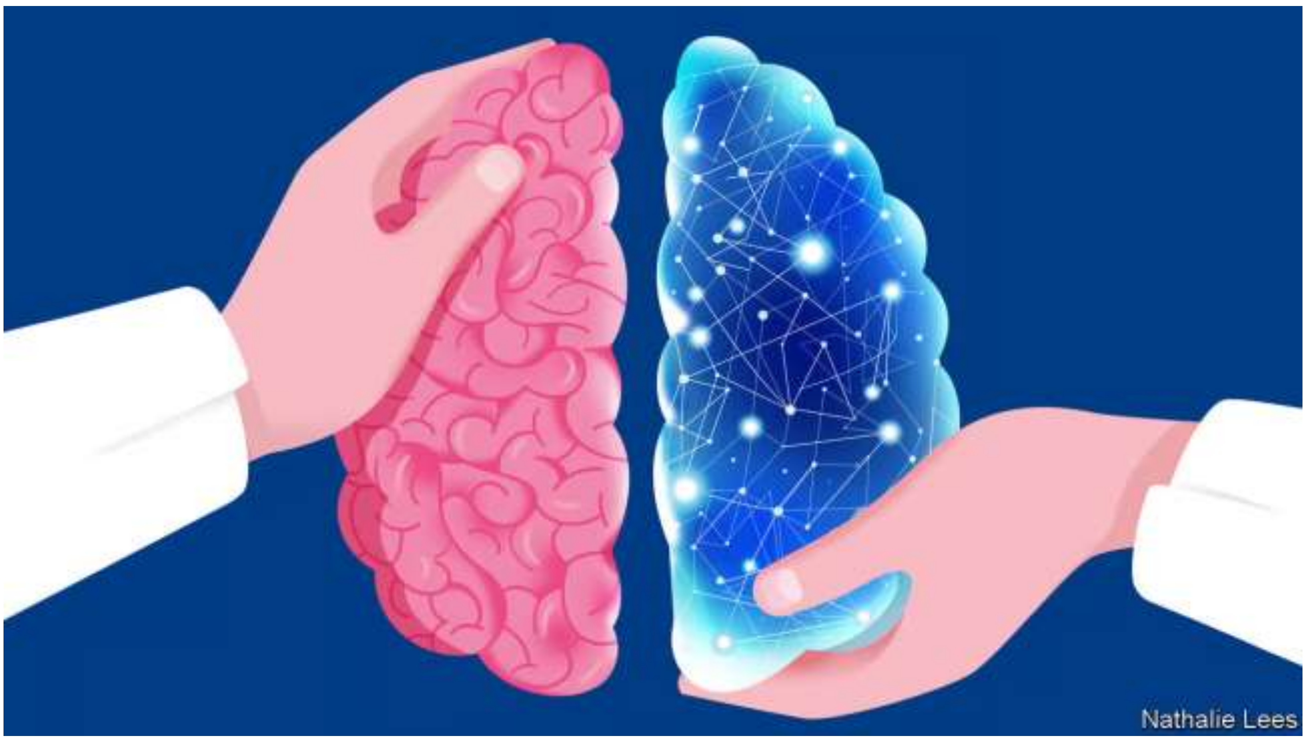
SAINTS PERSPECTIVES
Things Scientific and Technical

Artificial Brains are helping Scientists study the Real Thing

No model is perfect. But that doesn't stop them being useful

The Economist, May 24th 2023



The striking progress in artificial intelligence over the past decade is mostly down to advances in machine learning, whereby computers teach themselves complicated tasks by crunching large quantities of data, rather than having to be programmed directly by humans. This approach has driven rapid progress in computer vision, language translation and, most recently, the human-like conversational skills of **chatbots** such as **GPT-4**.

Listen to this story. Enjoy more audio and podcasts on iOS or Android.
The learning is done by software models called "artificial neural networks" (**ANNS**). The standard description of an **ANN** is that it is loosely inspired by the networks of neurons in the human brain. It is de rigueur to follow that

description with an immediate disclaimer, in which both computer scientists and neuroscientists jump in nervously to point out that **the analogy is very rough**, that ANNS are mere cartoons of real brains (if even that) and that they fail to capture the complexity of the biological organ.

All that is true. But some neuroscientists are beginning to find that even cartoons can be useful. The inner workings of the best ANNS—those that are closest to matching human performance on tasks like identifying objects, or responding to text prompts—appear to have some remarkable similarities to the workings of brains. Having taken inspiration from biology, in other words, programmers are now returning the favour, with their creations telling neuroscientists useful things about biological brains.

The seminal study comparing brains and ANNS was published in *Proceedings of the National Academy of Sciences* in 2014. Daniel Yamins, a neuroscientist at the Massachusetts Institute of Technology (MIT), and his colleagues trained an ANN to pick out objects from photographs—a cat, for instance. The researchers compared what was going on inside the electronic network to what was happening inside the brains of macaque monkeys that had been set the same task, and whose brains had been wired with electrodes.

ANNS are built up from large numbers of artificial neurons that, just like their natural counterparts, can be on or off; firing or silent. These neurons are linked together in layered, interconnected networks. Activity in lower layers can affect how neurons in the higher layers fire.

Inside the black box

Dr Yamins's test involves image recognition, which in natural brains proceeds hierarchically. One layer of neurons will detect simple features such as patches of light or dark. A higher layer organises those into edges; a still higher layer combines the edges into shapes. That process of increasing abstraction continues until, eventually, the brain decides whether it is looking at a cat, a dog or a banana.

Images that share some characteristics provoke similar clusters of neurons to fire. If a certain set of neurons fires when looking at a cat, another, partially overlapping set is likely to fire in response to a picture of a dog. The neurons that respond to both images are thought to be representing features—fur, four legs and a tail, say—that are present in both pictures.

When Dr Yamins and his colleagues compared what was going on inside the macaque brains to the silicon ones, they found arresting parallels between how the monkeys represented images and how the computers did. “The paper was a game-changer,” says Nancy Kanwisher, another professor at MIT who has spent much of her career studying the human visual system, and who now uses ANNs in some of her research. “The [artificial] network was not in any way designed to fit the brain. It was just designed to solve the problem and yet we see this incredible fit.”

Since then, whenever an ANN model has close to human performance on a task, neuroscientists have been eager to compare it with natural brains. They have found similarities between ANNs s trained to recognise speech and process language, such as those used in transcription software, and the human auditory cortex.

The pattern holds for written language too. One paper published in 2021 compared human brain activity against many different commercial language models. It found that the most sophisticated ANN—at the time **Openai’s** GPT-2—was the closest match for human brain activity. The better models get at solving certain tasks, the more similar they seem to get to the human brain doing the same.

Another indication that the analogy between artificial neural networks and natural ones is useful is that the study of the former can make testable predictions about the latter. A paper published in 2022, by researchers at Columbia University and MIT, found that an ANN trained on image-recognition tasks produced a group of artificial neurons devoted to classifying foodstuffs specifically. When the paper was published there was, as far as anyone knew, no analogous area of the human visual system. But the following year researchers from the same laboratory announced that they had discovered a region of the human brain that does indeed contain neurons that fire more often when a person is shown pictures of food.

Perhaps the strongest evidence for the claim that artificial brains can reveal useful things about biological ones is the apparent ability for software and wetware to interact with each other directly. Nicholas Sexton and Bradley Love, a pair of neuroscientists of University College London, started out rather **sceptical** about the supposed resemblance between natural and artificial neural networks. Simply seeing similar patterns of activity, they argued, was not enough to claim that ANNs and brains were solving problems in the same way.

To prove that the correspondence was meaningful, they suggested investigating whether it was possible to feed brain activity into an ANN.

In 2022 they published a paper in *Science Advances* that did just that. The researchers fed an ANN trained to recognise images data recorded by an MRI scanner examining human brains. The idea was to try to let the ANN "see" through human eyes. Sure enough, the hotwired ANN was able to interpret data from any of the hierarchical layers of the biological visual system—though it did best with data from the higher levels, which had already been partly processed by the brain in question. If the computer model was shown brain activity from a human that was looking at a picture of a greyhound, for example, then it would say that it was looking at a greyhound—as opposed to some other object—almost 70% of the time.

The fact that **a silicon brain can happily accept half-chewed data from a biological one suggests that, on some level, the two systems are performing the same sort of cognitive task**. That insight might prove useful for brain-computer interfaces, which are devices that aim to allow biological brains to talk directly to machines. An ANN linked up to a camera, for instance, might be used to feed partly processed visual information into the brain. That might help treat some forms of blindness caused by damage to the brain's visual system. Several different research groups in Europe and America are already testing that idea in experiments on macaques.

Models of the mind

Even those most enthusiastic about ANNs do not argue they are perfect analogues of the human brain. Some make mistakes that humans never would—give an ANN a picture of a cat but with the skin of an elephant, for example, and the model is more likely to identify it as an elephant. But no scientific model is ever perfect. The question is whether it is useful. One of neuroscience's problems is that experiments are difficult to run, for both ethical and practical reasons. Poking and prodding ANNs could offer a useful alternative.

In any case, comparing biology and silicon continues to produce intriguing results. In a paper published in May researchers from the *University of Texas at Austin* used a neural network to monitor brain signals from participants in an MRI scanner. Using just data from the MRI, the ANN could produce a rough summary of a story that the test subject was listening to, a description of a film they were watching, or the gist of a sentence they were imagining. “When I was in graduate school I would dream about something like this existing,” says Dr Love. “I thought it would be hundreds of years until we had something that works this well.” ■

Related Articles

Neural network - Wikipedia
https://en.wikipedia.org/wiki/Neural_network

Large, Creative AI Models will Transform Lives and Labour Markets
The Economist, Science & Technology, April 22nd 2023

Artificial intelligence - Wikipedia
https://en.wikipedia.org/wiki/Artificial_intelligence

Artificial intelligence Encyclopaedia Britanica
https://www.britannica.com/technology/artificial-intelligence/The-Turing-test