SAM FALCONER

ARTIFICIAL INTELLIGENCE

Marriage of mind and machine

Bringing together artificial intelligence and neuroscience promises to yield benefits for both fields.

BY NEIL SAVAGE

Chethan Pandarinath wants to enable people with paralysed limbs to reach out and grasp with a robotic arm as naturally as they would their own. To help him meet this goal, he has collected recordings of brain activity in people with paralysis. His hope, which is shared by many other researchers, is that he will be able to identify the patterns of electrical activity in neurons that correspond to a person's attempts to move their arm in a particular way, so that the instruction can then be fed to a prosthesis. Essentially, he wants to read their minds.

"It turns out, that's a really challenging problem," says Pandarinath, a biomedical engineer at Emory University and the Georgia Institute of Technology, both in Atlanta. "These signals from the brain — they're really complicated." In search of help, he turned to artificial intelligence (AI). He fed his brain-activity recordings to an artificial neural network, a computer

architecture that is inspired by the brain, and tasked it with learning how to reproduce the data.

The recordings came from a small subset of neurons in the brain — around 200 of the 10 million to 100 million neurons that are required for arm movement in humans. To make sense of such a small sample, the computer had to find the underlying structure of the data. This can be described by patterns that the researchers call latent factors, which control the overall behaviour of the recorded activity. The effort revealed the brain's temporal dynamics — the way that its pattern of neural activity changes from one moment to the next — thereby providing a more fine-grained set of instructions for arm movement than did previous methods. "Now, we can very precisely say, on an almost millisecond-by-millisecond basis, right now the animal is trying to move at this precise angle," Pandarinath explains. "That's exactly what we need to know to control a robotic arm."

His work is just one example of the growing interaction between AI and cognitive science. AI, with its ability to identify patterns in large, complex data sets, has seen remarkable successes in the past decade, in part by emulating how the brain performs certain computations. Artificial neural networks that are analogous to the networks of neurons that comprise the brain have given computers the ability to distinguish an image of a cat from one of a coconut, to spot pedestrians with enough accuracy to direct a self-driving car, and to recognize and respond to the spoken word. Now, cognitive science is beginning to benefit from the power of AI, both as a model for developing and testing ideas about how the brain performs computations, and as a tool for processing the complex data sets that researchers such as Pandarinath are producing. "The technology is coming full circle and being applied back to understand the brain," he says. That cycle of mutual reinforcement is likely to continue. As AI enables neuroscientists to obtain further insights into how computation works in the brain, the effort might lead to machines that can take on more human-like intelligence.

It's only natural that the two disciplines would fit together, says Maneesh Sahani, a theoretical neuroscientist and machine-learning researcher at the Gatsby Computational Neuroscience Unit at University College London. "We're effectively studying the same thing. In the one case, we're asking how to solve this learning problem mathematically so it can be implemented efficiently in a machine. In the other case, we're looking at the sole existing proof that it can be solved — which is the brain."

A BRAIN ANALOGUE
The successes of AI owe much to the arrival of more powerful processors and ever-growing quantities of training data. But the concept that underlies these advances is the artificial neural network. These networks consist of layers of nodes that are analogous to neurons. Nodes in the input layer are connected to nodes in a hidden layer by a series of mathematical weights that act like the synapses between neurons. The hidden layer is similarly connected to an output layer. Input data for a task such as facial recognition could be an array of numbers that describe each pixel in an image of a face in terms of where it falls on a 100-point scale from white to black, or whether it is red, green or blue. Data are fed in, the hidden layer then multiplies those values by the weights of the connections, and an answer comes out. To train the system to produce the correct answer, this output is compared with what it should have been if the output were an exact match for the input, and the difference is used to adjust the weights between the nodes. A more complex version of this process, called a deep neural network, has many hidden layers. It's this kind of system that London-based AI research company DeepMind Technologies, which is owned by Google's parent company, Alphabet, used to build the computer that beat a professional human player at the board game Go in 2015 — a victory widely hailed as a triumph for machine intelligence.

An artificial neural network is only a rough analogy of how the brain works, says David Sussillo, a computational neuroscientist with the Google Brain Team in San Francisco, California, who collaborated with Pandarinath on his work on latent factors. For instance, it models synapses as numbers in a matrix, when in reality they are complex pieces of biological machinery that use both chemical and electrical activity to send or terminate signals, and that interact with their neighbours in dynamic patterns. "You couldn't get further from the truth of what a synapse actually is than a single number in a matrix," Sussillo says.

"You couldn't get further from the truth of what a synapse actually is than a single number in a matrix."

Nonetheless, artificial neural networks have proved useful for studying the brain. If such a system can produce a pattern of neural activity that resembles the pattern that is recorded from the brain, scientists can examine how the system generates its output and then make inferences about how the brain does the same thing. This approach can be applied to any cognitive task of interest to neuroscientists, including processing an image. "If you can train a neural network to do it," says Sussillo, "then perhaps you can understand how that network functions, and then use that to understand the biological data."

DEALING WITH DATA
AI techniques come in handy not just for making models and generating ideas, but as a tool for handling data. "Neural data are terribly complicated, and so often we will be using techniques from machine learning simply in order to look for structure," Sahani says. Machine learning's main strength lies in recognizing patterns that might be too subtle or too buried in huge data sets for people to spot. Functional magnetic resonance imaging, for example, generates snapshots of activity throughout the brain at a resolution of 1–2 millimetres every second or so, potentially for hours. "The challenge of cognitive neuroscience is how you find the signal in images that are very, very large," says Nicholas Turk-Browne, a cognitive neuroscientist at Yale University in New Haven, Connecticut. Turk-Browne is leading one of several projects that are looking for fresh insights at the intersection of data science and neuroscience.

Using a machine to analyse these data is speeding up the research. "It's a huge change in how neuroscience is done," Sussillo says. "The grad students don't need to do as much sort of mindless work — they can focus on bigger questions. You can automate a lot of it, and you may get more accurate results."

REPRODUCING SENSES
Building an artificial system that would reproduce brain data was the approach taken by Daniel Yamins, a computational neuroscientist at the Wu Tsai Neurosciences Institute at Stanford University in California. In 2014, while Yamins was a postdoctoral researcher at the Massachusetts Institute of Technology in Cambridge, he and his colleagues trained a deep neural network to predict the brain activity of a monkey when it was recognizing certain objects (D. L. K. Yamins *et al. Proc. Natl Acad. Sci. USA* **111**, 8619–8624; 2014). Object recognition in humans and monkeys is performed by a brain system called the ventral visual stream, which has two main architectural features. First, it is retinotopic, which means that the visual-processing pathways in the brain are organized in a way that reflects how the eye takes in visual information. Second, the system is hierarchical; specific areas of the cortex perform increasingly complex tasks, from a layer that identifies only the outlines of objects to a higher one that recognizes a whole object, such as a car or a face. The details of how the higher layers work are poorly understood, but the result is that the brain can recognize an object in various positions and under different lighting conditions, when it seems bigger or smaller on the basis of its distance, and even when it is partially hidden. Computers are often flummoxed by such obstacles.

Yamins and his colleagues constructed their deep neural network according to the same retinotopic, hierarchical architecture as the brain and showed it thousands of images of 64 objects that varied in characteristics such as their size and position. As the network learnt to recognize the objects, it produced several possible patterns of neural activity. The researchers then

Computational neuroscientist Daniel Yamins is developing neural networks that can mimic brain activity.

compared these computer-generated patterns with patterns recorded from the neurons of monkeys while they performed a similar task. It turned out that the versions of the network that were best at recognizing objects were the ones with patterns of activity that most closely matched those of the monkey brain. "What you find is that the structure of the neurons is mimicked in the structure of the network," Yamins says. The researchers were able to match areas of their network to areas of the brain with about 70% accuracy.

The results confirmed that the architecture of the ventral visual stream is important for its processing ability. In 2018, Yamins and his colleagues performed a similar feat using the auditory cortex, in which they created a deep neural network that was able to identify words and genres of music from 2-second clips with the same accuracy as a human (A. J. E. Kell *et al. Neuron* **98**, 630–644; 2018). It helped researchers to identify which areas of the cortex perform speech recognition and which recognize music — a small step towards understanding the auditory system.

Neuroscientists are still a long way from understanding how the brain goes about a task such as distinguishing jazz from rock music, but machine learning does give them a way of constructing models with which to explore such questions. If researchers can design systems that perform similarly to the brain, Yamins says, their design can inform ideas about how the brain solves such tasks. That's important, because scientists often don't have a working hypothesis for how the brain operates. Making a machine perform a particular task will give them at least one possible explanation for how the brain achieves the same thing.

After researchers have built a hypothesis, the next step is to test it. Once again, AI models can help, by providing a representation of brain activity that can be tweaked to see which factors might be important in accomplishing a specific task. Researchers are limited by ethical considerations in terms of how much they can intervene in processes in the healthy human brain, so many recordings of neural activity in people come from the brains of those with epilepsy who are due to have brain tissue removed. This is because it is permissible to implant electrodes in brain tissue that will be excised anyway. Animal models enable researchers to use more invasive procedures, but there are human behaviours, notably speech, that cannot be replicated in other species. AI systems that can mimic human behaviour and be perturbed without ethical problems will provide scientists with extra tools for exploring how the brain works: researchers could teach a network to reproduce speech, and then impair that speech to observe what happens, for instance.

COMMON CONCERNS

Computer science and cognitive science are tackling some big questions, and working out how to answer them in either of these fields could drive both forwards. One such question is exactly how learning occurs. Neural networks mostly perform supervised learning. To master image recognition, for example, they might be shown images from ImageNet, a database of more than 14 million photographs of objects that have been categorized and annotated by people. The networks develop a statistical understanding of what images with the same label — 'cat', for instance — have in common. When shown a new image, the networks examine it for similar numerical attributes; if they find a match, they will declare the image to be that of a cat.

That's obviously not how babies learn, says Tomaso Poggio, a computational neuroscientist at the Center for Brains, Minds and Machines, which is part of the Massachusetts Institute of Technology. "A baby sees something on the order of a billion images in the first two years of life," he says. But few of these images are labelled — only a small proportion of objects will be actively pointed out and named. "We don't know how to deal with that," says Poggio. "We don't know how to have machines that learn from mostly unlabelled data."

His laboratory is in the initial stages of a project that would enable a neural network to perform unsupervised learning, by inferring patterns from unlabelled videos. "We know biology can do that," Poggio says. "The question is how."

Yamins is tackling unsupervised learning by devising programs that behave like babies at play, who interrogate their environment through random interactions and slowly develop an understanding of how the world works. He essentially codes in curiosity to motivate the computer to explore, in the hope that new behaviours will emerge.

Another outstanding question is whether some aspects of intelligence are 'installed' by evolution. For instance, people seem to be predisposed to recognizing a face as a face; babies can do so from the first hours of life. It might be, Poggio suggests, that our genes encode a mechanism for learning that task quickly and early in development. Deciphering whether that idea is correct could enable computer scientists to work out one way to help machines to learn. And other researchers are studying the neural basis of morality. "People are afraid of 'evil' machines," Poggio says. "We'd probably better know how our moral behaviour arises if we want to build good machines, ethical machines."

Yamins says that it is difficult to see how neuroscience alone will be able to uncover how unsupervised learning works. "If you don't have an AI solution, if you have nothing that works artificially, you can't possibly have a model of the brain," he says. It's more probable, he thinks, that computer scientists will come up with one or more solutions that neuroscientists can then test. "It might turn out that they're wrong," he says, "but that's why you check them out."

Answering these riddles could create more intelligent machines that are capable of learning from their environments and that can combine the speed and processing power of computers with more human abilities. The data-crunching and modelling abilities of computers are already bringing about advances in brain science that researchers say are likely to grow. "AI is going to have a huge impact on neuroscience," Sussillo says, "and I want to be a part of that." ∎

Neil Savage *is a science and technology journalist in Lowell, Massachusetts.*

CLARIFICATION
The Outlook article 'Marriage of mind and machine' (*Nature* **571**, S15–S17; 2019) omitted one of Chethan Pandarinath's affiliations: he is also at Emory University.