The Mind

Life & Arts

Can man ever build a mind?

Artificial intelligence has problems. The biggest is that brains are nothing like computers

Henry Marsh JANUARY 10, 2019

The idea that we might create machines more intelligent than ourselves is not new. Myths and folk stories abound with creations such as the bronze automaton Talos, who patrolled the island of Crete in Greek mythology. These stories reflect a deep, atavistic fear that there could be other minds that bear the same relationship to us as we do to the animals we eat or keep as pets. With the arrival of artificial intelligence, the idea has re-emerged with a vengeance.

We are condemned to understand new phenomena by analogy with things we already understand, just as the anatomists of old named parts of the brain after fruit and nuts — the amygdala (almonds) and the olives, to name but two. Although Hippocrates, in the fourth century BC, had firmly placed the brain at the centre of human thought and feeling, most early medical authorities had little use for it.

Aristotle thought it was a radiator for cooling the blood. The importance of the brain, for Galen 500 years later, were the fluid cavities — the ventricles — in its centre, and not the tissue of the organ itself. With the rise of scientific method in the 17th century, the brain started to be explained in terms of the latest modern technology. Descartes described the brain and nerves as a series of hydraulic mechanisms. In the 19th century the brain was explained in terms of steam engines and telephone exchanges.

And in the modern era, of course, the brain is seen as a computer. And yet, others argue, any metaphor is deficient: we have never met brains before and lack the appropriate language or concepts. The human brain, it is suggested, will never be able to understand itself. You cannot cut butter with a knife made of butter, as a neuroscience friend recently said to me.

Feedback

Henry Marsh © Tom Pilston

As a neurosurgeon, I have to accept that everything I am thinking and feeling as I write this is a physical phenomenon. Throughout my career I have seen patients who have suffered personality change (almost invariably for the worse) as a result of brain damage — it is difficult to believe in any kind of "mind" separate from the brain's matter when you see your fellow human beings changed, often grotesquely, in this way. And yet it is a deeply counter-intuitive thought. I remember one of my patients, when I was operating on his brain under local anaesthetic, looking at the computer monitor that showed the view of his brain down the operating microscopic that I was using. "This is the part of your brain which is talking to me at the moment," I told him, pointing with my instruments to the speech area of his left cerebral hemisphere. He was silent for a while, as he looked at his own brain. "It's crazy," he said.

The human brain, it is often stated, is the most complex structure in the known universe. It consists of some 85bn nerve cells, each of which is connected to many thousands of other nerve cells, with some 150tn connections. You can, however, if you wish, reduce it to something that

sounds quite simple. Each nerve cell is an input/output device that processes data. The input is electrical stimulation (or inhibition) by the thousands of other nerve cells whose output limbs, called axons, are connected to it via one-way connections called synapses.

The human brain, it is suggested, will never be able to understand itself. You cannot cut butter with a knife made of butter

Diagrams of brain nerve cells resemble trees in winter — the tangle of bare branches are the dendrites, where the synapses are found, and the tree trunk is the axon. The output is a train of electrical "spikes" fired down the axon, in response to the input of the nerve cells connected to it via the synapses. The axon is in turn connected to other nerve cells via the synapses on their dendrites. And so it goes, on and on, in a hugely elaborate dance. So, superficially, you might think the brain is a massive array of interconnected electrical switches. With modern, high-speed computers, which are massive arrays of interconnected electrical switches, surely it is only a matter of time before we build computers — artificial intelligences — that are smarter than we are? But how does such a simple design produce the extraordinary richness of human life and thought? And, for that matter, of animal life as well?

Recommended

Even our profound knowledge of physics and chemistry doesn't begin to explain how this electro-chemical activity generates conscious experience — the feeling of pain and the redness of the colour red, or the mysterious way that these seemingly identical electro-chemical processes produce sensations such as sound and vision that feel so different. And then there is "the binding problem": how does all this disparate neuronal activity, spread out in both space and time, produce coherent experience? With vision, for instance, we know that there are separate "centres" for colour and movement and other features such as vertical or horizontal lines; somehow, these bind and resonate together to produce a single image without, as one might naively expect, reporting in to some final integrating centre, synonymous with our sense of self.

The fact that the brain is a physical system must mean, of course, that it is subject to the laws of physics and is therefore computable. As the physicist David Deutsch has argued in his book *The Beginning of Infinity*, following on from Alan Turing's groundbreaking paper "On Computable Numbers", published in 1936, computers are not just another metaphor for the brain like steam engines or telephone exchanges. Computers are "universal simulators" and can, at least in theory, simulate all the information-processing that a human brain carries out, although it might take an almost infinitely long time. This is not to say that the brain is like a computer — a common misconception of Turing's position.

The fact that the brain must be subject to the laws of physics poses problems for the philosophers, who continue to ponder determinism and free will. Some, such as Daniel Dennett, end up claiming that consciousness and free will are illusions, a position that I, for one, find rather hard to apply to my day-to-day life, especially as I struggle to get out of bed on a cold winter morning. The more interesting questions are the practical ones — how close are computers to matching the human brain? How close are we to understanding how our brains work? And can AI help us understand our brains and vice versa?

―――

The complexity of the brain resides in the way in which the nerve cells are connected. Nerve cells are not, in fact, simple electrical switches. They are formed by organic molecules and not of the silicon and metal of computers. The connection at synapses is carried out by chemicals — neurotransmitters, of which there are upwards of a hundred different types. Neuronal networks are dynamic — they are constantly changing. They weaken or strengthen in accordance with how much traffic is passing through them. The brain is — to a certain extent — plastic. It can re-programme itself. In blind-folded people, for instance, cortical areas that normally process vision will start processing sound within 48 hours. The brain is topologically complex — it may consist of billions of similar brain cells and their connecting axons — but they are arranged in physically very distinct and precise groups and networks. Furthermore, neurons come in a wide variety of shapes and

sizes, even though they all share the same basic plan of dendrites, cell body and axon.
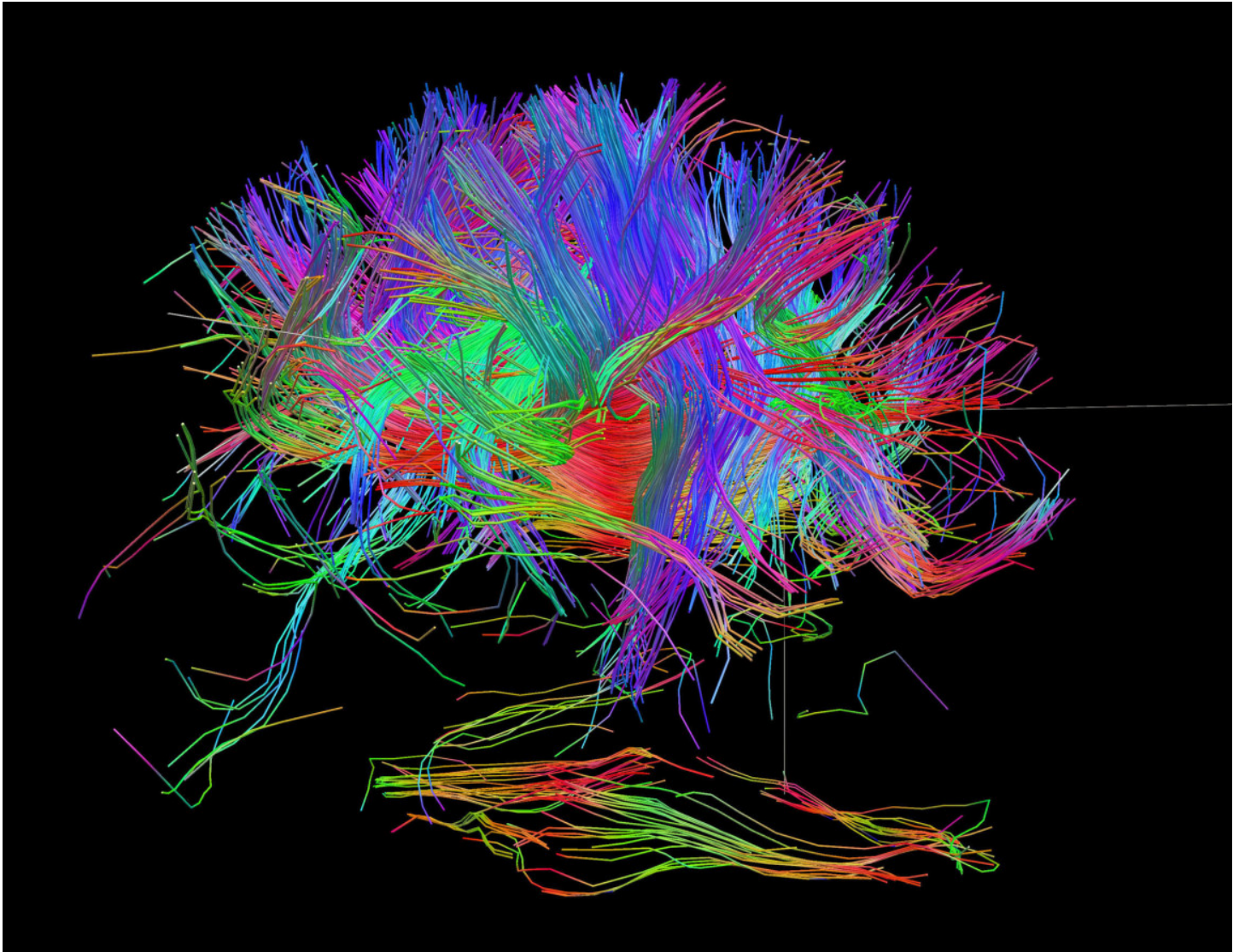

matter of the brain, produced by the Human Connectome Project, a multi-faculty collaboration that aims to produce a complete 3D wiring diagram of the organ © Courtesy of the USC Laboratory of Neuro Imaging and Athinoula A. omedical Imaging, Consortium of the Human Connectome Project, www.humanconnectomeproject.org

Brains do not come as isolated entities as do computers. They come with bodies, to which they are intimately connected. The bodies are their interface with the world, through which they learn to move, and, some would argue, there cannot be thinking without embodiment. The most plausible explanation for the evolution of brains is that they are prediction devices. To move effectively to find food, shelter and sex, we navigate using a model of the outside world created within our brains. Think of how confusing it is when you go down a staircase and there is one more or one less step than you unconsciously expected, or of the strange feeling as you reach the top of an escalator and the steps shrink. Perception is in large part expectation. Recent research is starting to reveal just how complicated the brain-body relationship can be — intestinal bacteria, for instance, seem to have a significant association with Parkinson's disease and autism — well, at least in mice. Human brains also come as part of social groups. We are, after all, utterly social creatures.

The philosopher David Hume wrote that "reason is, and ought only to be, the slave of the passions". Human intelligence is inextricably linked with emotions — the pleasure mathematicians find in an elegant proof, the curiosity that drives a young child to learn by exploring the world, the fear that helps us judge risk. The disembodied, isolated and emotionless "intelligence" of a computer is far removed from all this. This is not to say that true artificial intelligence — leaving aside the serious difficulty of defining the word "intelligence" in the first place — is impossible. It's

just that creating it might be rather difficult.
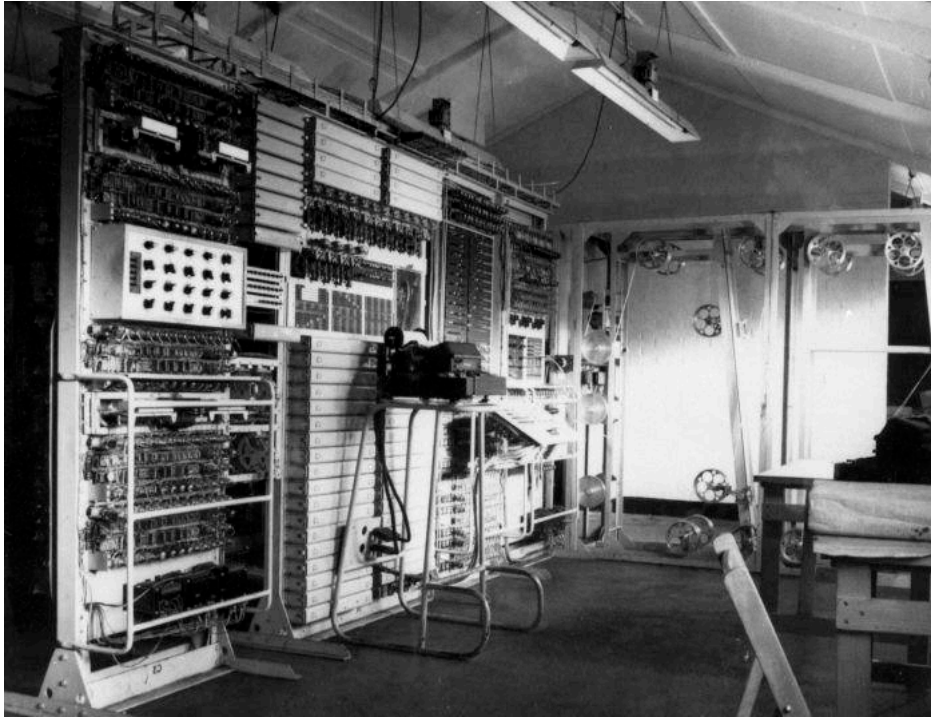
Colossus, the world's first electronic programmable computer, at Bletchley Park in Buckinghamshire. Bletchley was the British forces' intelligence centre during the second world war © Getty

There are significant practical limitations on the extent to which we can experiment and explore our own brains. The resolution of the best MRI brain scanners, for instance, is about one cubic millimetre, and one cubic millimetre of cerebral cortex can contain up to 100,000 neurons and a billion synapses. The temporal resolution is a little less than one second, and much cerebral activity is measured in milliseconds, so what we see with MRI scans are, in effect, blurred snapshots. Although higher resolutions are possible, they involve immensely strong magnetic fields with which there are many problems. Electrodes inserted into the brain can sometimes be used when operating on patients as part of their treatment work-up, and this can allow very limited experimentation, as can computer-brain interfaces implanted into paralysed patients who can then control, after much practice, robotic arms. But this is, so to speak, only scratching at the surface. We often have to use animals instead, and they can never tell us what they are thinking.

Brains do not come as isolated entities as do computers. They come with bodies, to which they are intimately connected

One of the many major research projects into the human brain is the US-led Human Connectome Project, a multi-faculty collaboration that aims to produce a complete 3D wiring diagram of the brain. This has recently been achieved, but only at intermediate resolution — not down to the level of individual cells and axons — for the mouse brain. It will take decades, not years, to produce the complete connectome of the human brain. One complete connectome has been established — that for the tapeworm *C. elegans*, which has some 307 neurons and 7,000 synapses. But even this took many years to unravel, and a connectome is only the beginning of trying to understand how a nervous system functions, just as looking at a map of a city is only the beginning of understanding a city. A connectome is necessary but not sufficient. Even when the human connectome has been created, ethical considerations may well limit the extent to which the connectome of dead brains can be related to the thoughts and feelings of living ones.

AI came into existence in the 1950s when the first digital computers were being built. At a famous — some would say infamous — conference at Dartmouth College in 1956, the early AI pioneers, such as Marvin Minsky, confidently asserted that machines would outstrip human intelligence within a few decades. But AI progressed unsteadily over the next four decades, with

several "AI winters", when funding dried up as a result of the signal failure of the field to deliver its promised results.

The early AI pioneer Marvin Minsky (centre) with other scientists at the Dartmouth Summer Research Project on Artificial Intelligence. The 1956 summer workshop is considered to be the seminal event for AI as a field © Courtesy of the Minsky family

The breakthrough in AI in recent years has resulted from the miniaturisation of computer chips. Computers have become more and more powerful, permitting massive and rapid data processing. It has also been made possible by abandoning attempts to programme computers with symbolic logic, which the early AI pioneers thought held the key to imitating human intelligence. The remarkable progress in AI in recent years is largely based on "neural networks" and machine learning, ideas developed decades ago but only recently put into practice. Neural networks only resemble brain networks in a very loose way. They consist of layered assemblies whose output can feed back and modify their input in accordance with a pre-programmed target, so that they "learn". They are engines of statistical association and classification. They neither explain nor do they understand. They have no internal model or theory of what is being analysed.

The literature, however, abounds in anthropomorphisms — AI programmes are said to "see" and "think". Google's DeepMind programme AlphaGo "vanquished" champion Go player Lee Se-dol. This is all nonsense. It is easy to get carried away. The predictive texting on your smartphone prompts you simply by calculating the probability of the next word from mindless analysis of previous text. Google Translate has trawled the entire contents of the internet without understanding a single word.
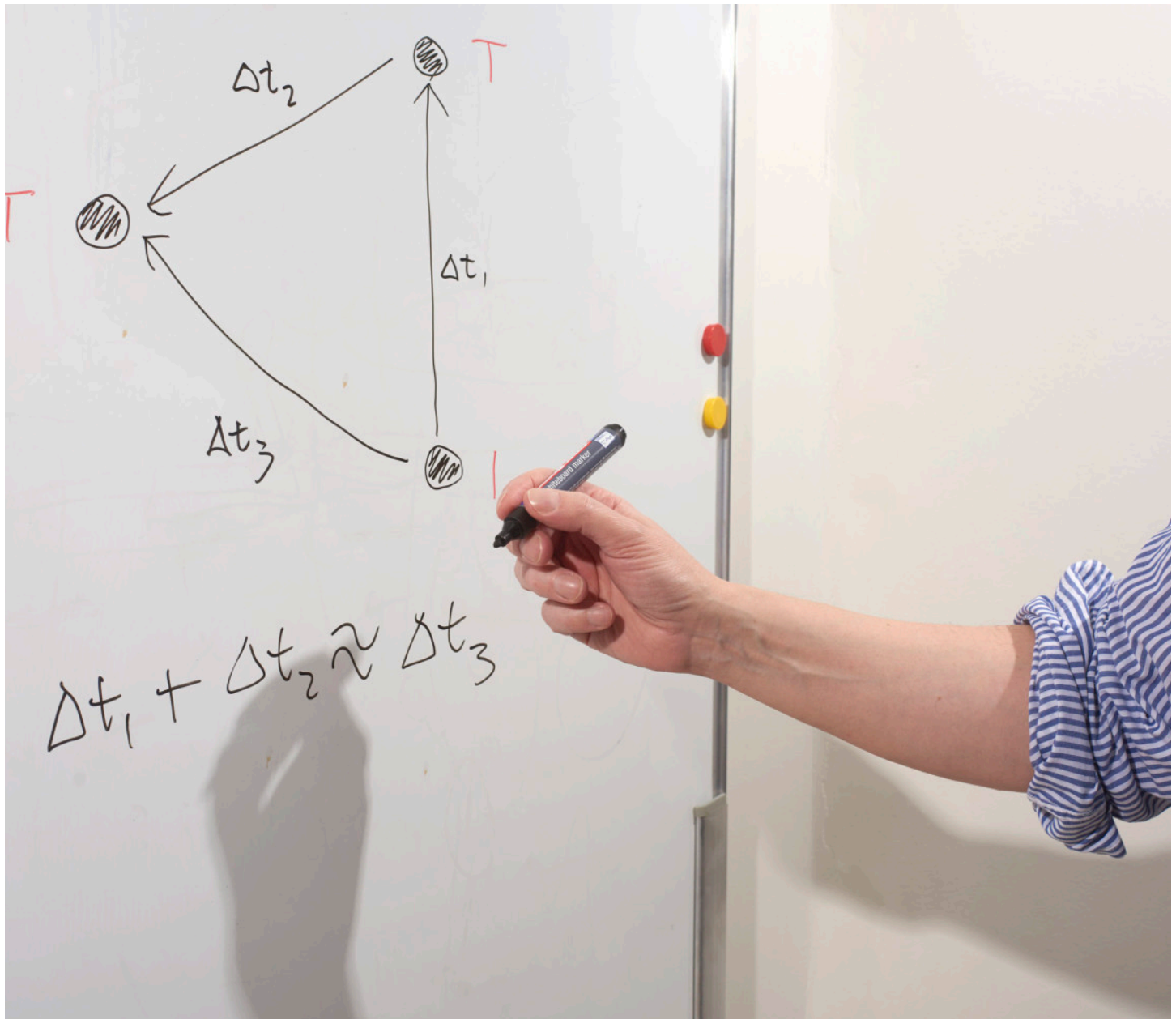
of the Oxford laboratory for Theoretical Neuroscience and Artificial Intelligence, explains the binding problem: how does disparate neuronal activity, spread out in both space and time, produce coherent experience? © Daniel Stier

This statistical approach to AI has nevertheless yielded remarkable results. And yet all current machine intelligences can only perform one task. This form of intelligence is reminiscent of the patients described in some of Oliver Sacks' writings — people who can carry out remarkable feats of calculation but are utterly helpless in normal society. And yet, despite the profound differences between the deep learning of AIs and organic brains, some recent research by the American neuroscientist Doris Tsao on facial recognition in monkeys shows that the facial recognition area in their brains uses algorithms very similar indeed to the ones used by AIs.

—

The Holy Grail for AI is "general intelligence" — a computer programme that could not only play games with simple rules but also perform other tasks, such as speech and face recognition, without being re-programmed. On present evidence it looks unlikely that neuronal networks and deep learning will ever be able to do this. Google recently created an AI that could recognise cats in photographs without even training the AI initially with pictures of cats. But this required the AI to be exposed to millions of pictures (it did not "look" at them), whereas a child probably only has to meet a few cats to be able to identify reliably all cats in future.

The energy consumption of a human brain is 20-30 watts — a dim lightbulb. An exascale computer, capable of a quintillion calculations per second, scaled up to the size of a human brain, would consume hundreds of megawatts. Computer engineers talk of the "von Neumann bottleneck", a problem with classical computer design in which memory is stored separately from the Central Processing Unit, and one of the reasons why computers use so much energy.
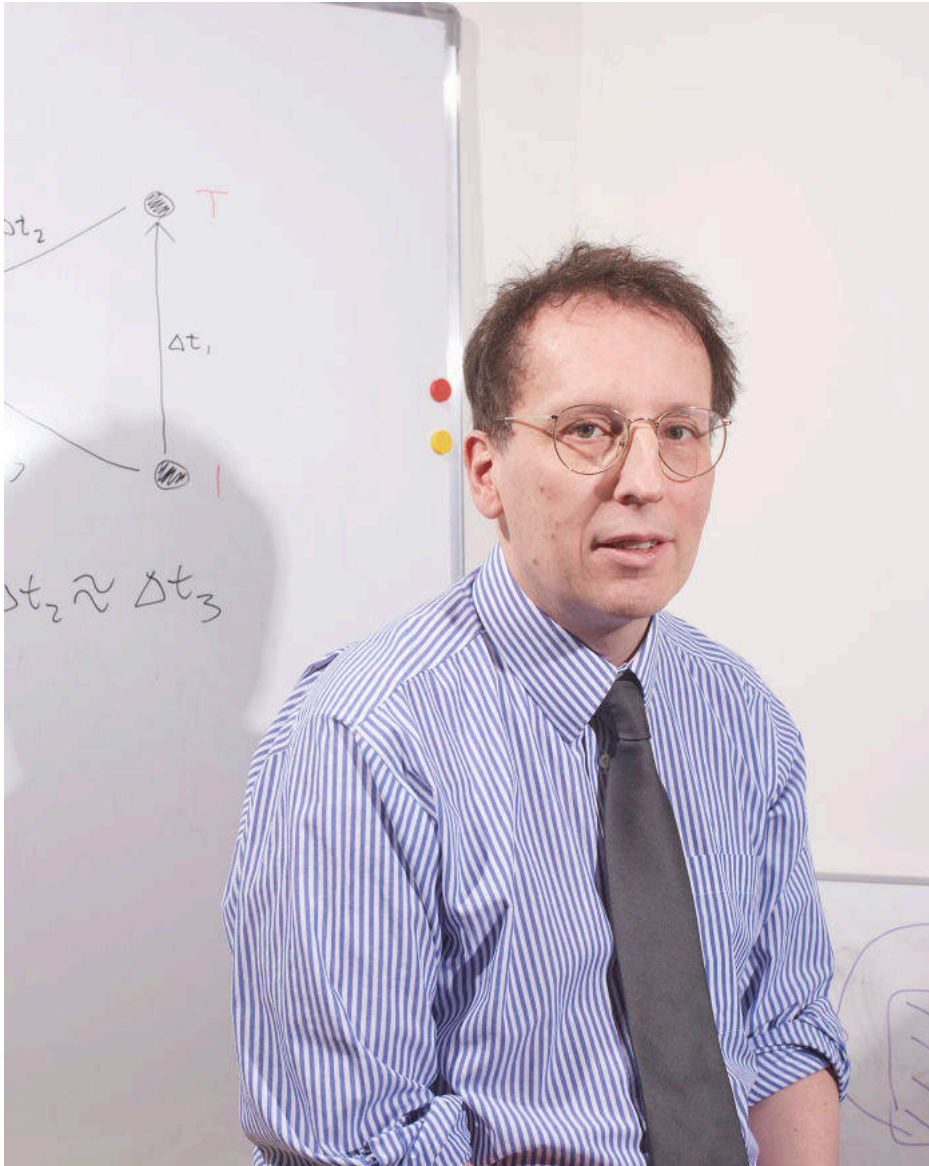
Dr Simon Stringer © Daniel Stier

Just as "neural networks" were inspired by biological brains, there is now great interest in "neuromorphic chips", where computer chips are designed to physically resemble nerve cells rather than merely mimic them in the software program. This should, in principle, make computers much less energy-intensive. The Human Brain Project (HBP), a multi-centre European research effort, funded by the EU to the tune of billions of euros, started with the grand ambition to build a brain — initially of a mouse, then of a human — from the bottom upwards, using computers. By creating a brain out of computer chips, the hope is that we might understand our own brains better. This is a somewhat optimistic inversion of Richard Feynman's famous remark: "What I cannot create, I do not understand." Whether the HBP will achieve this or not is highly controversial, given that our current understanding of how brains work is so limited.

Filled with a fervour reminiscent of the Second Coming and the Rapture, people talk of the

One part of the HBP, using neuromorphic design, is Professor Stephen Furber's experimental and newly launched SpiNNaker computer at Manchester University. It is designed to be both energy-efficient and to run models of brain function, among other things. It is too early to know, he tells me, whether it

Singularity, a time when AI will equal human intelligence

will lead to major breakthroughs or not, and this unpredictability is, of course, the very essence of science.

There are many other approaches. Simon Stringer and his team at the Oxford Laboratory for Theoretical Neuroscience and Artificial Intelligence have recently published a paper arguing that better theoretical models of neuronal transmission, in particular trying to account for the binding problem, hold more promise of progress. The timing of the firing of nerve cells, and the delays that occur in the time it takes the nerve impulse to travel along axons, could, they argue, be of critical importance in explaining the binding problem. There can be no doubt that much more fundamental research of this kind is required.

I find it strange that some people are so certain that the arrival of "superintelligent" machines is only a matter of time. Filled with a fervour reminiscent of the Second Coming and the Rapture, they talk of the Singularity, a time when AI will equal human intelligence. This belief — which is all it is — often comes with the hope that the human brain and all its contents can be re-written as computer code and that we will find the life everlasting as software programs. These ideas are not to be taken seriously, although they certainly sell books.

The future is always uncertain, and the future of AI and neuroscience is uncertain. Given the vast amount of current research in both fields, it is inevitable that tremendous progress will be made. Machine-learning AIs in their present form are already having a profound economic and social impact and will continue to do so. There is no risk, however, in the foreseeable future, of superintelligent AIs replacing us, or treating us as we have so often treated, and continue to treat, animals. Whether the effects of AI will be beneficial or malign will depend on the uses to which we ourselves put them. As to the possibility of machines acquiring general intelligence — "true AI" as some call it — it's anybody's guess. All that is certain is that even if it is possible, it is still a very long way away.

Henry Marsh is a neurosurgeon and author of 'Do No Harm: Stories of Life, Death and Brain Surgery' (Weidenfeld & Nicolson)

*Animated main image by Clara Bacou and Jessica Elian*

The Life & Arts team would love to hear from our readers. What do you think about recreating the human brain using AI? Do you think we'll ever get there? What are your specific fears?

Follow @FTLifeArts on Twitter to find out about our latest stories first. Subscribe to FT Life on YouTube for the latest FT Weekend videos

FT Insurance Innovation Summit

New York
11 April 2019

Combatting Inertia and Cutting Through the Noise
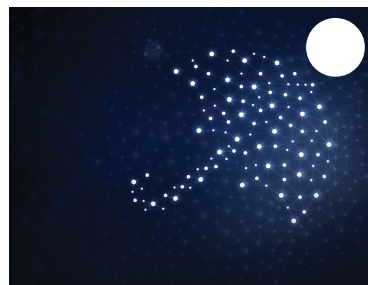
Register now   Presented by

Paid Post

Follow the topics in this article

Artificial intelligence

Science

Technology sector

Technology

Life & Arts