

## ▶▶▶ THINKING MACHINES

Artificial intelligence has changed how we interact with the world and with each other, yet actual thinking machines stubbornly refuse to exist. Will true AI ever be created, and would we even want it if it was, asks **Jason Walsh**

# THOUGHT EXPERIMENTS ON MAN, MIND AND MACHINE

**A**rtificial intelligence (AI) occupies a curious position in computing today, and even in wider society. AI applications, from chatbots to bank account spending analyses, are a part of everyday life, and yet the promise of true AI, of machines that can think for themselves, has never come to pass.

Although it has an extensive pre-history in myth and fiction, as well as early attempts by philosophers to represent human cognition in systematic terms, artificial intelligence had its genesis at Dartmouth College in 1956.

It was then that cognitive scientists Marvin Minsky and John McCarthy, along with mathematician Claude Shannon and computer scientist Nathaniel Rochester held the first ever AI workshop.

The workshop's proposal seems breathtaking now, but in an era that had seen mechanical computation devices give way to digital machines that we today would still recognise as computers, it

may have seemed reasonable to title the workshop "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."

Clearly confidence ran high. By 1967, Minsky had declared that the "problem of creating 'artificial intelligence' will substantially be solved" inside a generation. It wasn't. It still hasn't been.

Measured on its own terms then, AI has been a failure. And yet, both academic researchers and businesses have used AI to dramatically change how we live and work.

Professor Barry O'Sullivan, director of the Insight Centre for Data Analytics at the School of Computer Science and Information Technology at University College Cork, said the AI pioneers soon ran into hard problems.

"There was great optimism, but also a fundamental lack of understanding," he said. "There were technical, computational things they weren't aware of at the time. That optimism ran out of steam."

And run out of steam it did: in the 1980s the Japanese government poured billions into parallel processing research under the aegis of the Fifth Generation Computer Systems project. It went nowhere.

Elsewhere, a so-called AI winter saw researchers and businesses alike eschew the term AI in favour of elegant variations and euphemisms such as 'expert systems', knowing very well that by the mid-1980s, even using the term AI would see funding disappear in a puff of logic.

"As you try to improve it you come to appreciate how much more difficult it becomes, that's the story of AI," said O'Sullivan.

How can it be, then, that parallel processing is now commonplace and artificial intelligence is now a normal business operation? Just what

kind of failure is it that changes the world in such a manner?

Part of the answer is that as the field has matured it has better been able to define the questions it wants to ask. So rather than creating a simulated human, the focus is now on developing systems that can perform tasks once vouchsafed for humanity.

O'Sullivan said within AI research there was a distinction to be made between 'strong' AI and 'weak' or 'narrow' AI. Narrow AI performs tasks, as the name implies, in a narrowly defined area, but those tasks are nonetheless in areas usually associated with human intelligence.

"Strong AI is sometimes called artificial general intelligence [AGI]. Weak AI or narrow AI is the kind of applications we experience all the time, recognising faces or recommending things," he said.

And strong AI? As yet, it does not exist. "On the AGI side, there are people who think it is so far away that we shouldn't even worry about talking about it, but then you have the Elon Musks of the world talking about the existential threat to humanity. Most in the field think it is a fiction, and not even what the field is doing anymore," he said.

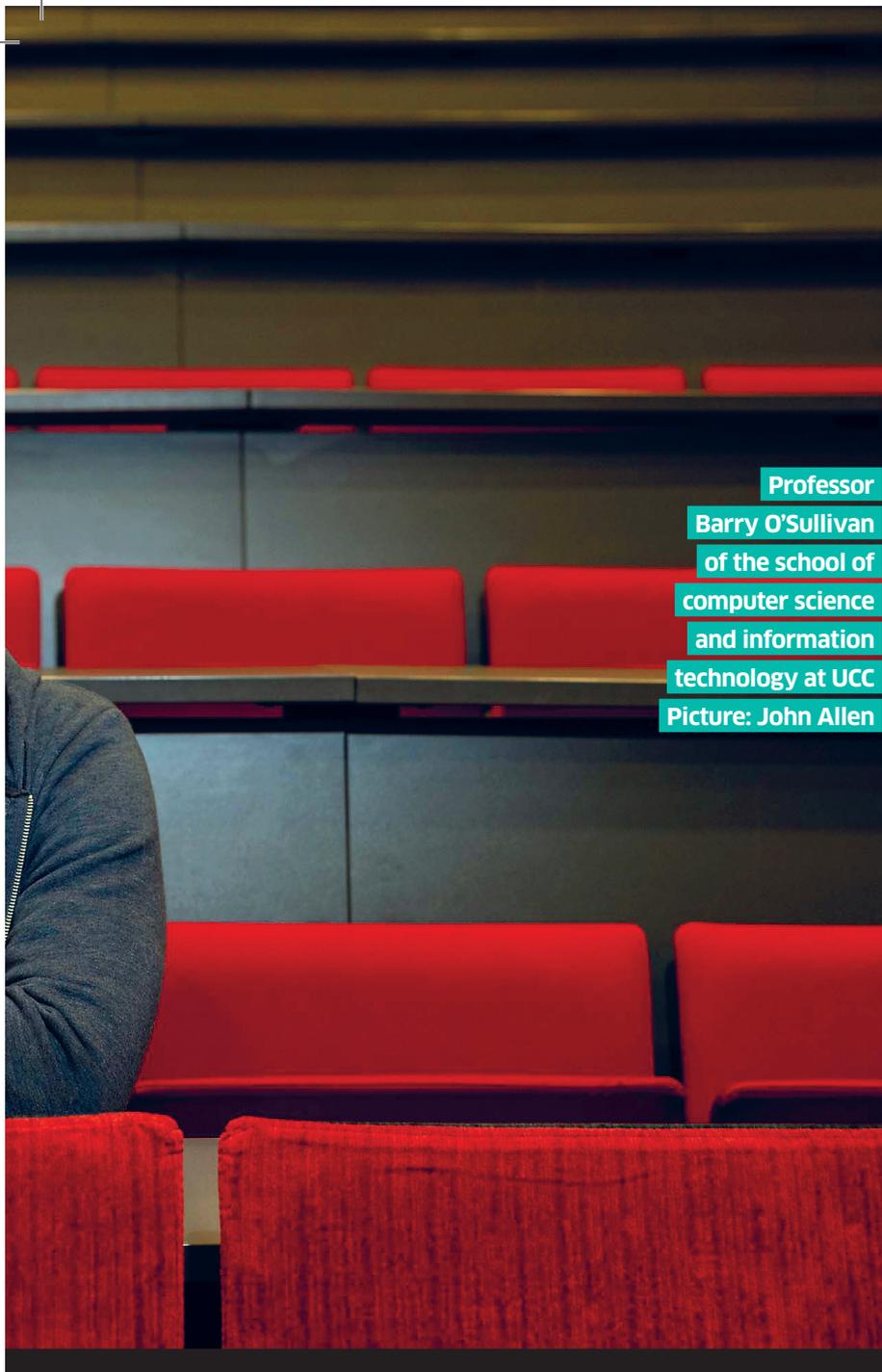
Today, bad press around AI tends to be focused on unwanted potential developments such as autonomous military machines and algorithmic bias.

Once a question for science fiction authors (for example, *Second Variety* by Philip K Dick examines the former, while *Cory Doctorow's Affordances* considers the latter) these are now real questions that need to be considered, and not only by scientists or even philosophers. Today, machine intelligence is a question for politicians and, ultimately, citizens.

Fears about the impact of technology are hardly new, though, including how information collection and processing can be used against people.



**Dr Stephen Rainey,**  
a research fellow  
at the University of  
Oxford's Uehiro Centre  
for Practical Ethics



Professor  
Barry O'Sullivan  
of the school of  
computer science  
and information  
technology at UCC  
Picture: John Allen



*A computer remains a tool and just as a hammer cannot put a nail in a wall – the doing is in the agency of the holder of the hammer – a computer does not itself truly perform the tasks we use it for*

“I remember reading a paper from the 1980s on how harmful this technology would be in terms of bias, and that technology wasn't AI, it was databases,” said O'Sullivan.

### Is AI impossible?

Even if killer robots are banned – though it seems unlikely they will be, as they represent a perfect example of the prisoner's dilemma: if one country bans them, then it is left at the mercy of those which don't – and algorithmic bias corrected through legislation and better data, there remain other questions to be asked about AI. Indeed, AI raises questions about ourselves.

Dr Stephen Rainey, a research fellow at the University of Oxford's Uehiro Centre for Practical Ethics said philosophers like him were interested in AI as it related to fundamental questions about humanity.

“It raises questions about knowledge in general. If a computer can know something, well, how does it know that and what does it mean for us knowing things?” he said.

In practical terms, one question might be: if a computer can fill in a form better than you can, if we as humans copy the computer can we become better at it?

More broadly, among those who think about AI there are two camps: those who think it is all about data, and those who say intelligence is qualitatively different and not just an outcome of operations being performed on data. This division cuts to the core of an ancient question: what is it to be human?

“There are epistemological questions and they have knock-on effects for philosophy of mind,” said Rainey.

But can a machine be said to actually know

anything? There are reasons to think not. Rainey said he was not convinced that a machine following steps, even steps that were inferred rather than explicitly programmed, qualified as thinking.

One key issue is that humans think in complex ways, with ideas processed through both reason and effect. Even if a machine can be said to reason, it has no effect – which means any potential thinking machine would qualify as a psychopath.

“Reasons and emotions are related things for us, but for a computer it's not the case. For a computer it's all about weighted probabilities. We can think like that, but it's not the only way we think,” Rainey said.

There is even the question of whether a computer can actually be said to be doing anything, let alone thinking. Obviously, in a straightforward sense a computer is capable of doing things: accepting input and providing output to the user, for example. More fundamentally this is a representation of simple computational mathematics, so something is certainly happening.

And yet, a computer remains a tool and just as a hammer cannot put a nail in a wall – the doing is in the agency of the holder of the hammer – a computer does not itself truly perform the tasks we use it for.

The existence of autonomous vehicles transforms what might otherwise seem an abstruse thought experiment into a practical question, one with major ethical implications.

“If you have an AI driving a car there is no control, the AI has no interests. You could argue that the AI isn't driving the car because driving implies a whole series of things,” Rainey said.

“What is it that distinguishes you deciding to roll down a hill on a sunny day for fun from you falling down a hill? The distinguishing factor is the reason, and a computer has never had a reason, it only has the next step.”

Human reason and effect, though, are not the only forms of thought we experience, and the existence of inference implies that machines can be taught a lot more.

For example, if you go outside your house and you recognise a cat, you don't need to reason what a cat is and that you see it. Advanced uses of AI today tend to fall into this category.

“Recognising diseases, discovering things, translating languages, all of these things are perceptive tasks,” O'Sullivan said.

“If you were being cynical about it, you could say that what goes on in the human brain is that we learn a complicated function, and this function has millions or billions of inputs, and because of the availability of hardware we can now do that.”

However, the availability of data, storage capacity and computational power, which has undoubtedly driven advances in AI, arguably more so than have new fundamental breakthroughs, does not mean that truly intelligent machines will spring forth any time soon.

“One of the fundamental challenges in AI is that many of these systems don't have common sense, and common sense in AI has a formal meaning. ML systems that are based on data don't have that common sense, they lack a causal model and that's actually going to be a big problem because people will expect common sense reasoning,” O'Sullivan said.

### Future tense

As developments continue, politicians and the public alike are forced to grapple with the question of AI more in terms of ‘what do we want from it?’ than ‘do we want it at all?’

In a world where there are more truck drivers than computer programmers – which is to say, despite appearances, the world we actually live in – the spectre of economic devastation cannot be brushed aside.

O'Sullivan said that instances of unfairness deriving from technology like AI rarely came from people being made unemployed.

## Thinking machines

The history of artificial intelligence is characterised by waves of hope, hype and failure, but the tide seemed to turn with the ‘expert systems’ of the 1980s and, more dramatically, the rise of cloud computing in the 2010s.

However, with many a start-up today claiming to have AI at the heart of its operations, it is useful to have some working definitions.

The AI imagined in the 1950s – machines that could actually reason – does not exist and when it is spoken of it is typically now referred to as artificial general intelligence.

The most common form of AI today is machine learning (ML), a subset of the wider field that uses algorithms to have machines automatically improve through experience.

Typical applications include chatbots, spam filtering and anti-malware software that learns how network traffic typically functions and can spot deviations from the norm.

A more recent subfield is deep learning (DL), which seeks to advance machine learning by modelling operations on the human brain, specifically neural networks.

Although this sounds, and indeed is, close to the Dartmouth definition, the goal is still not the creation of actual artificial reasoning so much as increasing the abilities of ML systems to make accurate predictions and use ever-larger datasets.

“The issue really is, are people getting a fair share of the benefits of the technology. If you look at manufacturing over the last 50 or 60 years, it has been highly automated,” he said.

“There hasn't been a catastrophic fall in employment numbers, but there has been wage stagnation and people have become net poorer.”

In the event that AGI was ever developed, however, there would at least still be work for philosophers.

“How would you know if there was AGI? Imagine an AGI candidate robot and its task was to make a cup of tea in your house: it has to make all sorts of subtle inferences to just get into your house and work out where the kitchen is,” said Rainey.

According to Rainey, such a seemingly novel artefact sounds suspiciously like something we are all already familiar with. “What you end up talking about is a human,” he said.

Whatever ethical implications that might throw up, it at least clarifies one way of thinking about ourselves: we are beings that exist phenomenologically.

“Human beings are essentially interpreters of things. Most AI is simulation behaviour. I'm more inclined to believe simulation behaviour exists than artificial intelligence,” he said.

O'Sullivan said essential questions did need to be asked but, as often as not, the most pressing questions were about the actions of humans, not machines.

“One researcher said that there's nothing artificial or intelligent about artificial intelligence. I gave a talk a while back, saying it's not artificial intelligence we need to worry about, it's real stupidity,” he said.