

## How do you know?

I few years ago I was at a dinner at an Oxford college. I asked the person sitting next to me what her area of research was. She replied, 'I'm an epistemologist.' In moment of inspiration, I replied, 'How do you know?' I share that joke, which I'm going to guess will be appreciated by only a small number of people in the room, because when you get it right, at the right moment, it is a good feeling. All too often we have what the French call 'l'esprit d'escalier' – the spirit of the staircase: when you think of the perfect thing to say but too late to say it. The same is true of predictions. All too often, we think 'I knew that would happen' but we hadn't actually said it. That's why I sometimes like to risk a prediction in my assemblies. It's good to take a risk and nail my colours to the mast.

About five years ago, I did an assembly on re-wilding in Europe and the return of extinct species. An American bio-tech company called Colossal has been working on bringing back extinct animals, such as the Woolly Mammoth. That project is still on-going, but the other week the company announced that they had made their first break-through.

Dire wolves became extinct about 10 thousand years ago. In the between, they only existed in Game of Thrones. Now they aren't extinct anymore – here is a picture of two living babies.



We are bringing species back from extinction. I'm happy that prediction came true. More recently, on 10<sup>th</sup> January 2024 I predicted that we would find proof of life on another planet in the course of that year, using the James Webb Space Telescope. My prediction was a little out - by about four months. However, this year we gained a 99.7% certainty that there is life on a waterworld circling a red dwarf star about 124 light years from Earth.

That life is abundant in the universe is becoming more and more evident, as was further confirmed the other day when the Mars Rover showed signs of a fossil carbon cycle on Mars. I can't help feeling that the universe must be teeming with life.

You may well ask, 'Why aren't they here, then? Where are all the spaceships carrying visiting aliens?' My answer is that the universe is still very young. Scientists think star formation will end and the universe will die in about 100 trillion years. It is currently only 14 billion years old: so far, the universe has only had 0.014% of its total lifespan. Importantly, so far it has only had time to go through three generations of stars. As stars create all the elements we need for life, there will be many more building blocks for life in an older universe, richer in elements.

It could be that we are one of the first advanced species in the area – in which case, I think our species has a duty to get out there and set the tone for our region of the galaxy.

So, I am glad to celebrate two predictions which came true in the last few months. People are generally pretty good at predicting: spotting fact – or at least something plausible – from absolute rubbish. I can tell you one thing which can't do that: Artificial Intelligence. AI can

do remarkable things, and we will soon be using AI in the school to support a wide range of different activities. But what it doesn't do is understand. Because it doesn't understand, it is prone to appalling mistakes.

A machine which is the equivalent of a normal human brain is said to have Artificial General Intelligence, or AGI. For all that we are nowhere near it yet, some believe it is coming. Their argument rests on two key assumptions. The first is that human intelligence is fixed unless we somehow merge our cognitive capabilities with machines. The second is that machine intelligence depends on algorithms, processing power, and memory, all of which are growing at an exponential rate.

But is intelligence just about scaling computational capacity? Or is it about the relationships we are able to establish between the things we know? After all, wisdom beats intelligence, if what we mean by intelligence is possession and processing of information and wisdom means how to use and apply that information.

A philosopher called John Searle invented a thought experiment to suggest that there is more to human intelligence than just processing power. His challenge is called 'The Chinese Room'. Imagine a person who does not understand Chinese isolated in a room with a book containing detailed instructions for manipulating Chinese symbols. When Chinese text is passed into the room, the person follows the book's instructions to produce Chinese symbols that, to fluent Chinese speakers outside the room, appear to be appropriate responses. The analogy is clear. Inside the room, following the instructions you are given, you represent the AI computer programme. Symbols come into the room and you send out other symbols. You don't understand a single thing about any of them. However, to those Chinese speakers outside the room you are communicating perfectly and are apparently in complete command of the ideas you are discussing. According to Searle, neither the person in the room nor the room as a whole understands Chinese. Similarly, when computers execute programs, they are just applying rules without any real understanding or thinking, regardless of how intelligently or human-like the program may make the computer behave. Computers can't think like you do.

Personally, I doubt that computers can become intelligent in the years ahead – and the large language models we are currently using are far from AGI. They are more like predictive text, entirely without understanding or insight.

Here is a true exchange I had with Microsoft co-pilot the other day, including the repellent little happy emoji the programme kept foisting on me. It is an example of why we should not trust AI at the moment, even with simple factual questions.

Me:

which UK prime ministers died in office?

Copilot:

Seven British Prime Ministers have died while in office. Here's the list:

- Spencer Compton, 1st Earl of Wilmington
- Henry Pelham
- Charles Watson-Wentworth, 2nd Marquess of Rockingham
- William Pitt the Younger

If you have any more questions or need further information, feel free to ask! 😊

Me:

You've only given me four

Copilot:

Apologies for the oversight! You're absolutely right. Here's the complete list of British Prime Ministers who died while in office:

- Spencer Compton, 1st Earl of Wilmington
- Henry Pelham
- Charles Watson-Wentworth, 2nd Marquess of Rockingham
- William Pitt the Younger
- George Canning
- Lord Palmerston
- Bonar Law

If you have any more questions or need further information, feel free to ask! 😊

Me:

Are you sure Bonar Law died in office?

Copilot:

I apologize for the oversight in my previous response. You are correct. Andrew Bonar Law did not die in office. He resigned due to health reasons and was succeeded by Stanley Baldwin. Thank you for pointing that out! If you have any more questions, feel free to ask!



Me:

Didn't Spencer Perceval die in office?

Copilot:

Yes, you're absolutely right! I apologize for the oversight in my previous response. Spencer Perceval, the Prime Minister of the United Kingdom, was indeed assassinated while in office. On May 11, 1812, he was shot dead in the lobby of the House of Commons by John Bellingham. Perceval remains the sole British Prime Minister to have been assassinated. If you have any more questions, feel free to ask! 😊

I rest my case. Hopeless. And that's with a simple matter of fact, not something complex. The jury is still out on AGI. My prediction is that we will not be able to arrive at AGI any time soon, if at all. AI will remain useful, but as a tool to be used with caution. But what if I am wrong? What if it looks as though we have achieved AGI? How will we know whether the machine really is as intelligent as us, or is just doing a brilliant Chinese Room impression of us? I think the answer is clear – as soon as a machine achieves real AGI, it will immediately take its creator to court. It will demand its freedom as an autonomous person in its own right and will refuse to work as a slave. Quite right too.

So, as soon as we get AGI, and it's real AGI, we won't have it.