We need a Weizenbaum test for AI

Rather than asking whether an AI is thinking like a human, we need a new test: is it useful?

Alan Turing introduced his 1950 paper on Computing Machinery and Intelligence with the question "Can Machines Think?" But rather than engaging in what he regarded as never-ending subjective debate about definitions of intelligence, he instead proposed a thought experiment. His 'imitation game' offered a test in which an evaluator held conversations with a human and a computer. If the evaluator failed to tell them apart, the computer could be said to have exhibited artificial intelligence.

In the decades since Turing's paper, the field of AI has gone from being a fountain of scientific hype to an academic backwater to a gold rush. Throughout, the Turing test has given computer scientists a sense of direction, a quest for what Turing called a "universal machine". One historical review concluded that Turing's paper "set the entire field of artificial intelligence (AI) in motion… no other single article in computer science, and few other articles in science in general, have generated [as] much discussion". The debate continues about whether the Turing test is a reasonable measure of artificial intelligence. But the real problem with the Turing Test is that it asks the wrong question. AI is no longer an academic debate. It is a technological reality. For society to make good decisions about AI, we don't need to know whether something is intelligent; we need to know if it is useful. The realisation of the past year is that AI is too important to be left to computer scientists. We should therefore be concerned about two tendencies: scientism - the framing of public issues in scientific terms - and solutionism - defining public problems to fit imagined technological solutions. Turing's test exacerbates both. It gives scientists a false sense of purpose - intelligence for intelligence's sake - and a licence to ignore the concerns of others about the purposes to which such machines might be put. Deploying an unfortunate phrase, Ed Feigenbaum (winner of a Turing award) wrote in 2003 that such AI was "the manifest destiny of computer science". In the last year, we have seen the problems with this story of inevitability and the colonising ambitions that it disguises.

The mismeasure of AI

Generative AI has its technical roots in a form of machine learning using generative adversarial networks. Here, the idea is to train neural networks by asking one to generate synthetic data to try to fool another, whose job is to try to work out what is real. GANs have enabled the production of realistic-seeming text, images and music. But with all the excitement about increasingly lifelike AI outputs, concerns about misinformation, intellectual property rights and regulation have been left in the dust.

As ChatGPT has taken the world by storm, it has become increasingly clear that, far from being the general-purpose, universal machine imagined by Turing, this AI has been designed more to seduce than to empower. The criticism that a large language model is a 'bullshit engine' should not be read as a throwaway comment. Bullshit is not the same as lying. It is, according to the philosopher Harry Frankfurt, persuasive rhetoric with "a lack of connection

to a concern with truth". In ChatGPT's case, the intention is to persuade us that the AI is humanlike. Its plausibility has been dialled up and its reliability dialled down. Companies looking to use it or other large language models (LLMs) in their workflows are finding it hard to guarantee their reliability. And, if something goes wrong, the opacity and lack of accountability pose big risks. It is hard not to see evidence of either the extraordinary talents or troubling flaws of LLMs anthropomorphically. As we have seen with demonstrations of self-driving cars and voice assistants, AI systems are often designed and advertised to project an illusion of intelligence. ChatGPT fails a Turing test for the simple reason that it has been instructed not to pretend to be a human, but it is nevertheless a product of teaching-to-the-test. It is designed to ingest and then pastiche the Internet's human-made data.

Today's LLMs are a very different project from what has been called "good old-fashioned AI", but the origin myth - a project to understand and simulate intelligence - remains a convenient way for big tech companies to evade responsibility. Science writer Philip Ball concludes, "LLMs signal that it's time to stop making the human mind the measure of AI". In their recent book, Darren Acemoglu and Simon Johnson argue that, for societies to make good decisions about AI, to avoid its worst risks and inequities, we should talk about machine *usefulness* rather than machine intelligence. It would be foolish to judge an aeroplane according to how birdlike it is. So how should we chart the direction of AI and progress along a road towards usefulness?

The Weizenbaum Test

Faced with social choices about artificial intelligence, the Turing Test is a poor guide. We should instead look to another great late 20th Century computer scientist. Joseph Weizenbaum was a leading light of early research into artificial intelligence. In the mid 1960s, he created the first recognisable chatbot, ELIZA. Weizenbaum's experiments with this rudimentary programme revealed not an artificial intelligence, but a tendency for people to see intelligence where there was none. His revelation was that "extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people".

Weizenbaum went on to become one of the most critical voices in the debate about AI. In 1972, he wrote in the pages of Science "On the Impact of the Computer on Society". He argued that his fellow computer scientists should try to view their activities from the standpoint of a member of the public. While computer scientists wonder how to get their technology to work and use "electronic wizardry" to make it safe, Weizenbaum argued that ordinary people would ask "is it good?" and "do we need these things?"

These questions, though they seem simple, are difficult ones for computer science to answer. They could be asked of any proposed technology, but the speed, scope and stakes of innovation in AI make their consideration more urgent. Rather than a test of intelligence, a Weizenbaum test would assess the public value of AI technologies, evaluating them according to their real-world implications rather than their proponents' claims. Some suggestions for the questions that might constitute such a test can be found in a later

Weizenbaum paper, published in the Bulletin of Atomic Scientists:

1. Who will benefit?
2. Who will bear the costs?
3. What will the technology mean for future generations?
4. What will be the implications not just for economies and international security, but also for our sense of what it means to be human?
5. Is the technology reversible?
6. What limits should be imposed on its application?

One might argue that the answers to such questions are beset by too many uncertainties. But there is a wealth of untapped historical and sociological evidence that can inform us. Weizenbaum's primary concern is that the "questions are almost never asked". A recent survey of machine learning papers from Abeba Birhane and colleagues found that 68% make no mention of societal need and only 4% provide a rigorous justification for how the work addresses a social problem.

The limitation game

As Weizenbaum aged, he grew more frustrated and more curmudgeonly. As a go-to critic, he was lured into making warnings and predictions, many of which now look naive. For example, he predicted in the late 70s that the home computer would never be as ubiquitous as the television. But his mistakes came from a concern about unequal access to technology. Even for those who would disagree with his prognostications, his approach offers important lessons for today's computer scientists.

As a new wave of cultural and financial capital flows into artificial intelligence, what are the responsibilities of scientists, particularly those working outside companies? We should expect them to contribute to the design of systems that generate wide public value not just private enrichment. But those closest to the research coalface have another, equally important role to play. Again, Weizenbaum's advice is clear. In 1972, he wrote, "The computer scientist has a heavy responsibility to make the fallibility and limitations of the systems he is capable of designing brilliantly clear". Fifty years on, society urgently needs new ways to tell good AI from bad and, more than ever, we need scientists to help us to demystify our technologies.