Can machines think?

Subhashis Banerjee [*]
(Views expressed are personal)

March 29, 2024

Can machines think? The popular understanding seems to be that the day is not far when an Artificial Intelligence (AI) will be able to think like humans and interact – at least through languages – in a way that is indistinguishable from real humans. The coming of such a day has been termed as "the singularity", a pivotal moment for the human race. With the recent success of the Large Language Models (LLM) like ChatGPT with language interpretation and composition, many think that the day is imminent.

When confronted with the possibility, Ludwig Wittgenstein, one of the most influential philosophers of the 20^{th} century, famously said that "But a machine surely cannot think!" He probably meant that the concepts of thinking and intelligence can only apply to living objects; it would be grammatically and logically incorrect otherwise! Nevertheless, machines can indeed share some traits of human behaviour, so, even without precise definitions of these terms, their increasing use for machines is perhaps germane. In fact, in the eventuality that we do go past the said "singularity" – a proposition that sounds frightening – a machine may someday have to be treated like a "person"!

The universal computer

Most folks trained in computer science must believe that such AI must be possible. This is because central to the accepted theory of computation – as obtained from Alan Turing in 1936 among others – is the existence of an abstract algorithmic concept of a universal computer, that can simulate the actions of all other computers. At a slight cost of over-simplification, one can think of this universal computer as one that can execute any program written in any modern programming language given unbounded memory and time. Of course, it may not be able to do so "efficiently", but that is only because we may not yet have discovered a sufficiently efficient model of computation. Given adequate time and memory, the universal computer can in principle simulate, with arbitrary precision, all physical and chemical processes of the brain and other parts of the human body, and actually all of mother nature's, provided their theories are understood. David Deutsch, a physicist, philosopher and computer scientist who first formulated quantum computation, calls this a fundamental law of physics and computer science.

Of course, Turing fully understood universality and believed that AI must be possible. If so, it will also need sensorimotor perception, because it cannot possibly rely on external intelligence to provide it with the essential methods to internally survive and exchange signals with the outside world. Turing also estimated that the resources required to simulate a human brain, which he argued must also be a universal computer, will not be very large – in fact, less than that of a typical modern laptop. After all, the size of a human brain is not all that much. And, the fact that there must

---

[*] Professor of Computer Science, Ashoka University. Also with the Centre for Digitalisation, AI and Society. Email: suban@ashoka.edu.in

exist computational problems that cannot be solved by a universal computer – as established by the Gödel's incompleteness theorem and Turing's own results on computability – did not deter his arguments. After all, even intelligent human computers cannot solve many problems.

He also formulated a test for AI that a human judge should be unable to tell whether interacting with a human or a program. Many believe that current state-of-the art LLM-based AI software like ChatGPT built using deep neural networks may have come close to passing the *Turing test*.

What's the right theory of intelligence?

So, do we know how the brain works to be able to program a universal simulator for AI? Can a parametrised neural network model with parameters estimated using a purely data-driven inductive method be a program for the universal simulator? Unfortunately, the answers to these have to be a resounding no! We are nowhere even close.

Indeed, as this example demonstrates, ChatGPT does not seem to understand even the basic logic of numbers, despite its copious training from examples. In general, logical deductions – perhaps like most other cognitive tasks – cannot be extrapolated or generalised or inductively derived purely from data, which is what current state-of-the-art AI systems are based on. For example, no amount of training data can give us a mathematical abstraction like the Pythagorean theorem; it had to be deduced logically using created representations like numbers. And even with logical deductions there is a fundamental computational resource limitation problem. We know from theory of computation that most logical deductions are computationally intractable, and that there are an infinite hierarchy of logical deduction problems whose solutions will require ever increasing time and memory resources. Both induction and deduction - the two main methods of computation – are limited, and we clearly do not have the right theory of intelligence as yet.

A stone, a watch and a frog

Also, scientific theories are not read from observations in nature. They are obtained through a process of abduction, by making hypotheses – sometimes with wild guesses – and critiquing and reasoning about them, often with physical experiments but not always. Indeed, we have obtained fantastic theories like quantum mechanics, and gravitation based on curved space-time, only using such methods. They were only validated post-facto with observational data.

Moreover, despite its obvious appeal, the Turing test is inadequate for intelligence. It requires a judge to empirically decide whether an AI is indistinguishable from humans. However, judging a genuine AI will invariably require explanations of how it works. A purely behavioural test is bound to be insufficient because it is well known even in probability theory that, in general, multiple, possibly infinite, internal configurations and explanations of systems will exist that can result in same behavioural manifestations over the observables. It is like a philosopher trying to tell a living object by just looking at a stone, a watch and a frog. Also. the Turing test does not reveal who was responsible for the AI's behaviour. If it was an external human designer, then the program is not an AI.

Can machines think?

Nevertheless, it is the quest for passing the test that has brought the AI systems where they are. They indeed are impressive in their conversational coherence and there can certainly be many engineering applications where they can be used effectively. That will however require that conventional safety principles of engineering are adhered to. In that sense, the Turing test has certainly been remarkably useful.

Programming intelligence will thus require us to cross new epistemological barriers. Pure empiricism and inductive reasoning from data, using fake-it-till-you-make-it type optimisation, or even logical deductions, cannot possibly be adequate theories of intelligence. And, we don't even know how to algorithmically make wild guesses and hypotheses, let alone critiquing and analysing them. We are also fairly clueless algorithmically about emotions, and feelings like pain and happiness, and of course about sensorimotor perceptions.

The ultimate test of AI will have to be based on explanatory theories about it. And, if we understand them, then we should be able to program them. We have to reluctantly admit, however, that if ever we discover a theory of AI, it is more likely to emerge from the discipline of philosophy than from computer science.