What Is Artificial Intelligence?

Brendan Balcerak Jackson

A science of possibilities

As a field of research, artificial intelligence (or AI) is *the scientific study of intelligence in artificial systems*. But what does this actually mean?

To start with, we can think of an "artificial" system as any non-biological system designed by humans. These days, virtually all AI research is done with electronic computers. But the idea of AI is much older than computers. The seventeenth century philosopher Gottfried Wilhelm von Leibniz (1646-1716) speculated about the possibility of reducing all human reasoning to mechanical operations; to test his ideas, he designed the Stepped Reckoner, a hand-cranked mechanical device that could solve arithmetic problems.[1] Even as far back as the ancient Greeks, stories were told of how Hephaestus, the Greek god of fire, forged mechanical handmaidens out of gold who could speak with him and assist him with his work. This is one of the earliest appearances of the ideas of AI and robotics, long before digital computers were ever conceived.

As simple as this definition of "artificial" is, it highlights one of the most distinctive and exciting things about AI as a science. In many branches of science, the goal is to make discoveries about the world around us and how it works. Astrophysicists look for principles that govern the formation of galaxies, for example, and biologists look for the evolutionary forces that lead to new species. By contrast, AI scientists are not trying to discover how things are out there in the world as it already is. They are exploring *possibilities* for designing and building new things. Artificial intelligence is a science that studies what we can create.

Two kinds of intelligence

So much for "artificial". What do we mean by "intelligence"? This is a bit trickier. Modern AI researchers like to make a distinction between two senses of intelligence. Sometimes, when we describe a creature or system as intelligent this is because it is capable of reason and reflection. It uses the information at its disposal to think its way through some problem or question. Sherlock Holmes is a great example of intelligence in this sense. So are you, when you solve a tricky logic puzzle or work through a difficult a math problem. But you are also using intelligence, in this sense, when you do something as simple as keep mental track of how long your bread has been in the toaster in order to know when it is ready. By contrast, the toaster itself is not intelligent; it probably

[1] Beeson (2004)

just uses a simple bimetal-strip thermostat to "know" when the toast is ready. Let us call intelligence in this sense *intelligence-as-thinking*.

At other times, our interest in intelligence is in *action* rather than thinking – in the ability of the creature or system to make deliberate choices about how to behave in different situations. This is the sort of intelligence an expert chess player uses when she plays a game of chess. Of course, a lot of thinking goes into the way she plays the game. But it is the moves themselves – her actions – that make her an intelligent player. You demonstrate the same sort of intelligence even when you do something as simple as decide where to put your chair in the garden to get the best sun. When a sunflower turns to face the sun, by contrast, it is not acting intelligently; its behavior is just the result of an adaptive bio-chemical process. We can call intelligence in this sense *intelligence-in-action*.

Let's look at examples of modern AI research on each of these two kinds of intelligence.

Acting like a human: the Turing Test

One very famous example of work on intelligence-in-action in artificial systems is the Turing Test, which was first proposed by one of the pioneers of modern AI, the mathematician and computer scientist Alan Turing (1912-1954).[2] The Turing Test is a method for determining when an artificial system qualifies as intelligent.

A simple version of the test begins with one human, the *examiner*, sitting at a terminal that is connected to two other terminals. At one of these terminals sits another human, and at the other is the artificial system being evaluated for intelligence, the *test system*. At the start of the test, the examiner does not know which terminal leads to the human and which to the artificial system. Her task is to try to figure this out, by holding text conversations with each of them via her terminal. The task of her conversation partners is to make this as hard as possible for her, by doing their best to participate in the conversation just as an ordinary human would. A test system passes the Turing Test when it is reliably able to fool the examiner, which we can confirm by observing that the examiner's accuracy rate at identifying the artificial system is no better than random guessing (50%).

Clearly, the Turing Test is designed to detect intelligence-in-action rather than intelligence-as-thinking. To pass the test, there is no specific kind of reasoning or information-processing the system needs to go through. Whatever it does internally, the Turing Test measures the intelligence of the system entirely by its replies to the messages of the examiner – that is to say, by its actions. This might make it seem like the Turing Test poses a fairly straightforward challenge for AI. But in fact, it turns out to be a remarkably hard challenge to meet. In 1990 began an annual competition called the Loebner Prize, with a top prize for the first system that could definitively pass the Turing Test. To date, the top prize has never been awarded.[3]

[2] Turing (1950)

[3] `https://aisb.org.uk/`

Thinking like a human: the Winograd schema

Many contemporary AI researchers are dissatisfied with the Turing Test as a way to determine the intelligence of an artificial system. One reason is that systems can be designed to use silly jokes and conversational tricks which help mislead the examiner, but don't necessarily have much to do with real intelligence. One alternative that has been proposed uses the *Winograd Schema*, which are pairs of sentences of a certain type.[4] Here is an example:

(1) The town council members refused to grant the angry demonstrators a permit because they feared violence.

(2) The town council members refused to grant the angry demonstrators a permit because they advocated violence.

A typical English-speaking reader is very likely to interpret the pronoun *they* differently in the two sentences. She is likely to understand sentence (1) as saying that it was the town council members who feared violence, and to understand (2) as saying that it was the demonstrators who advocated violence.

These interpretations are not forced on us by the rules of English grammar. Those rules leave it open whether *they* refers to the council members or to the demonstrators in both sentences. We arrive at our interpretations by engaging in a very quick bit of implicit reasoning. We tend to assume that angry demonstrators are more likely to advocate violence than council members, and that town councils might consider the risk of violence when deciding whether to award protest permits. We combine this knowledge with our knowledge of grammar to arrive at the conclusion that the pronoun *they* must refer to the town council members in sentence (1), and to the demonstrators in sentence (2).

If an artificial system were consistently able to arrive at the same interpretations of sentences like (1) and (2) as we do, this would be a sign that the system was actually thinking about language like we do, rather than merely behaving like we do, as with the Turing Test. This is the idea behind the Winograd Schema test. The test system is given a large number of sentence pairs, and we check whether it is able extract the same information from them that an ordinary human English speaker would. When the system is given sentences (1) and (2), for example, can it figure out who was being said to fear violence, and who was being said to advocate it? The Winograd Schema test is clearly a test for artificial intelligence-as-thinking. Of course, the system must do *something* to let us know what answers it has figured out. It might display its answers on a screen, for example, or enter them into a database. But what matters for the test is that it figures out the answers, not that it behaves in any particular way. The Winograd Schema test is certainly no less demanding than the Turing Test, and it is perhaps not surprising that no artificial system has yet managed to come very close to human-level performance on it.[5]

[4] Levesque (2014)

[5] Larson (2021)

Human intelligence and rationality

There is another distinction to be made, one that cuts across the distinction between intelligence-as-thinking and intelligence-in-action. We ordinary humans are certainly intelligent creatures, in both of these senses. But we are hardly perfect. We sometimes make mistakes in our reasoning, we overlook relevant evidence or jump to conclusions, or we get overwhelmed by too much information. And we sometimes act impulsively or in ways that are not in our own best interests, or we fail to consider other things we could have done that would have been better. In short, we sometimes think and act *irrationally*.

It is natural to wonder whether we can build artificial systems that do better, systems that are free of our flaws and limitations. Such fully rational systems would have many obvious benefits. So, while the goal of developing human-like intelligence in artificial systems is certainly worthwhile, an equally worthwhile goal is developing systems that come closer to the ideal of fully rational intelligence.

The Turing Test and the Winograd Schema test are examples of human-like AI: both are tests for determining whether an artificial system has the kind of intelligence that we ordinary humans have. But examples of work on rational AI are easy to find too.

Acting rationally: autonomous vehicles

One of the most prominent contemporary examples of rational AI is work on the development of self-driving cars and other autonomous vehicles. Part of the appeal of self-driving cars, of course, is convenience: instead of having to drive we can read or take a nap and let the car do the driving. Beyond that, however, autonomous vehicles offer the promise of *better* driving, driving that reduces accidents and traffic congestion while increasing energy efficiency. Fulfilling that promises requires the development of AI that is more rational than us – in fact, it requires the development of several distinct rational AI systems.

For example, for a self-driving car to be better than an ordinary human at avoiding accidents, it needs to have a system that can identify objects in the environment faster and more accurately than we do, in order to spot potential hazards. (Is that a plastic bag blowing across the road ahead, or a loose dog?) And once a hazard is identified, it needs to have a system that makes faster and smarter decisions than we do about how to respond. (Is it better to swerve than to brake in this case, and if so in which direction?) If the vehicle is to reduce congestion and increase efficiency, it needs to have a route-planning system that is better than we are at factoring in all kinds of relevant and changing information – time of day, weather, ongoing construction projects, and so on – and it needs to have a system that makes more rational decisions about how fast or slow to drive along the way. The goal of autonomous vehicle research is to develop these and other AI systems, and to put them together to make a vehicle that is a more intelligent driver than ordinary humans.

Thinking rationally: using AI to predict protein structures

Proteins plays a central role in most biological processes, and every protein folds into a complex three-dimensional shape that determines the way it functions. Understanding the shapes of proteins is thus vital for making advances in medicine, epidemiology, and many other areas. Unfortunately, the fold structures of proteins are extraordinarily complex, and even when you know the amino acids that make up a particular protein, it is extremely difficult to predict how it will fold. This means that the fold structure of each protein needs to be discovered independently, using highly detailed and difficult laboratory procedures. So, while scientists have identified millions and millions of different proteins, they have so far managed to determine the fold structures of only a small fraction of them.

To address this problem, in 1994 AI researchers began holding a biannual competition called CASP, which stands for "Critical Assessment of Protein Structure Prediction."[6] AI systems that enter the competition are given the amino acid sequences for 100 different proteins, and the challenge is to work out the fold structure of each; the system with the highest accuracy wins. To date, the best performer in the CASP competition is AlphaFold, a machine-learning program that achieves an impressive 93% accuracy level, which is in fact slightly better than the traditional laboratory methods. [cite]

The AI systems that enter the competition are clear examples of artificial intelligence-as-thinking. Their purpose is to take in information about amino acid sequences and make inferences about fold structure. What actions these inferences lead to, if any, are not determined by the systems themselves. But the aim is not to design systems that think the way ordinary humans do, as in the case of the Winograd Schema test. Rather, the aim is to develop systems that are not bound by the cognitive limitations that we ordinary humans run up against when we grapple with the enormous complexity of protein structures. The goal is to develop systems that are better at thinking about protein structures than we are, systems that are in that sense more rational thinkers than us.

Summary

We have seen that modern research in AI divides into four distinct areas or branches. One branch focuses on intelligence-in-action, and is concerned primarily with decision-making and behavior in artificial systems. Another branch focuses on intelligence-as-thinking, on reasoning, and the manipulation of information in artificial systems. And in both of these areas, one branch of research is focused on developing human-like intelligence – systems that think or act like us – while another branch aims to develop systems that are more fully or ideally rational. In all four of its branches, AI is a science of possibilities: it investigates possibilities for creating forms of intelligence other than ourselves.

[6] `https://predictioncenter.org`

References

Beeson, M. J. (2004). The mechanization of mathematics. In *Alan Turing: Life and legacy of a great thinker*, pages 77–134. Springer.

Larson, E. J. (2021). *The Myth of Artificial Intelligence: Why Computers Can't Think the Way We Do*. Harvard University Press.

Levesque, H. J. (2014). On our best behaviour. *Artificial Intelligence*, 212:27–35.

Turing, A. (1950). Computing machinery and intelligence. *Mind*, LIX(236):433–460.