
What is artificial intelligence and what can it do?

JOSEP DOMINGO-FERRER

Institut d'Estudis Catalans
Universitat Rovira i Virgili
josep.domingo@urv.cat

Introduction

According to Boden (2018), “artificial intelligence seeks to make computers do the sorts of things that minds can do”. This definition has two problems. First, it does not explore the nature of intelligence. Secondly, it seems to cover two things that are not necessarily related to intelligence. So, the question is what kind of intelligence does artificial intelligence (AI) have? According to Bonnay (2022), intelligence is associated with successfully doing certain types of things, that is, doing them in a certain way, and something in us enables us to do them in that way.

Back in the seventeenth century, there were disputes as to whether or not intelligence was exclusive to humans. Descartes (1637) claimed that an automaton could not do what an intelligent being knows how to do. However, Hobbes (1651) said that reasoning is nothing other than calculating, which could open the door to an artificial kind of reasoning.

The history of AI

The term *artificial intelligence* was coined at a scientific gathering held at Dartmouth College (USA) in the summer of 1956, whose participants included Shannon, Minsky and other fathers of computing. Its underpinnings include Charles Babbage and Ada Lovelace’s analytical machine (1834), McCulloch and Pitts’s artificial neuron model (1943) and Turing’s machine and the Turing test (1950).

From the very start, there have been two distinct views of AI. Symbolists believe information is represented by symbols and their relations. In their view, AI consists of manipulating symbols

to solve problems. In contrast, connectionists do not believe information is represented explicitly but instead seek to imitate natural biological systems through artificial neural networks.

After the initial enthusiasm about AI's possibilities waned, there were few practical advances in the 1960s (called *the first AI winter*). During the 1970s and 1980s, symbolism yielded tangible results such as the languages Lisp and Prolog, as well as expert systems (such as the celebrated MYCIN, which reproduced doctors' reasoning to make diagnoses). The 1990s were another period of stagnation (*the second AI winter*). Since 2000, however, connectionism has clearly predominated thanks to the training of artificial deep neural networks with massive data and graphic processors.

Does AI think?

Turing (1950) proposed a test to determine whether a machine behaved intelligently. Basically, if person A cannot distinguish between the written responses given by computer B and person C, computer B could be said to be *functionally* intelligent.

Searle (1980) presented the argument of the Chinese room to refute the fact that computers can think. The idea is that a person who doesn't know Chinese is put inside a locked room. Messages written in Chinese are given to them through a hole in the wall. The person has a rulebook that enables them to use the incoming Chinese messages to produce outgoing messages in Chinese that make sense. The messages they produce are then sent out through another hole in the wall. Even though functionally the person inside the room seems to know Chinese, the reality is that they do not but instead are merely manipulating symbols. The Chinese room is a metaphor for a computer programme, which also manipulates symbols. The conclusion is that *functioning* as intelligent can be done with no understanding. Therefore, Turing's functional definition of intelligence has problems.

What do machines know how to do?

Machines, meaning computers equipped with artificial intelligence, can do a variety of tasks.

- Prediction. This category includes classification (character recognition, giving credit, etc.) and medical diagnosis (both the kind done by expert systems like MYCIN and automatic image diagnostics).
- Translation and transcription. These capacities are now available to the public at large (Google Translate, Word transcription and others).

- Strategy games. In 1997, the computer Deep Blue beat Kasparov as the world chess champion. In 2011, the computer Watson won the Jeopardy championship. In 2017, the computer Alpha Zero learned autonomously until it became the chess, shogi and go champion. However, these computers have limitations in their reasoning. For example, Watson would say that a crocodile cannot jump over obstacles because it has not seen one do that on the Internet, while a person would reason that it cannot because it has short legs and is clumsy (Levesque, 2014).
- Content creation. Here we can distinguish between the creation of artistic contents made with generative models (generative artificial networks, GAN) and the generation of discourse made with the main language models (ChatGPT and others). There is a philosophical debate on whether or not machines can be creative. Kant (1790) said that “fine art is only possible as the product of genius [...] It is a talent for producing that for which no definite rule can be given”. Kraus (1986), however, viewed creativity as derivation and recomposition, meaning that it could be automated.

The main language models

BERT (Google AI, November 2018) is a pre-trained language model that provides numerical representations for phrases and words which can be refined for specific tasks. Embeddings (or numerical representations) are vectors based on the fact that words that have similar meanings occur in similar contexts. Thus, they represent meaning as co-occurrence: the more co-occurrence there is between two words or phrases, the less distance there is between the vectors that represent them. This procedure is based on the work of linguists; more than 60 years ago, Firth (1957) claimed that “you shall know a word by the company it keeps”.

ChatGPT or Chat Generative Pre-trained Transformer (Open AI, 2022) is a conversational robot based on a massive language model. It has been trained with all the content on the Internet, especially Wikipedia. It can explain concepts using simple words, give ideas for birthday gifts or make computer programmes. In a conversation, it remembers what the user said before and introduces any corrections that the user tells it to correct. However, sometimes it can provide inaccurate, harmful or biased information. Researchers at Microsoft claim that ChatGPT shows “sparks” of general artificial intelligence (Bubeck *et al.*, 2023). Indeed, ChatGPT can do a wide range of tasks competently, including generating images, programming computers, applying mathematical skills, interacting with the world and interacting with humans.

Conclusions

AI has a history full of ups and downs. The current plethora of data and graphic processors has made it possible to train deep neural networks, so the discipline is currently in a sweet spot. Conversational robots have made AI visible to the public at large, although for years now it has been used for many other purposes (translating, choosing contents, etc.). Nonetheless, we do not seem to be close to general artificial intelligence, which would resemble human intelligence.

References

- BODEN, Margaret A. (2018). *Artificial Intelligence: A Very Short Introduction*. Oxford: Oxford University Press.
- BONNAY, Denis (2022). "Que peuvent les machines ? Petite philosophie de l'IA par exemple". In *Philosophie du langage ou de la logique* [online]. Description of the course at Université de Paris-Nanterre. <https://formations.parisnanterre.fr/plugins/odf-web/odf/_content/course-philosophie-du-langage-ou-de-la-logique-fr/Philosophie%20du%20langage%20ou%20de%20la%20logique%20.pdf> [Retrieved: 1 October 2024].
- BUBECK, Sébastien [et al.] (2023). *Sparks of Artificial General Intelligence: Early Experiments with GPT-4* [online]. <<https://arxiv.org/pdf/2303.12712.pdf>> [Retrieved: 22 November 2024].
- DESCARTES, René (2000 [1637]). *Discours de la méthode*. Paris: Flammarion.
- FIRTH, John R. (1957). "A Synopsis of Linguistic Theory, 1930-1955". In *Studies in Linguistic Analysis*. Oxford: Blackwell, pp. 10-32.
- HOBBS, Thomas (2017 [1651]). *Leviathan*. London: Penguin Classics.
- KANT, Immanuel (2015 [1790]). *Kritik der Urteilskraft*. Cologne: Anaconda Verlag.
- KRAUS, Rosalind E. (1986). *The Originality of the Avant-Garde and Other Modernist Myths*. Cambridge, Massachusetts: The MIT Press.
- LEVESQUE, Hector J. (2014). "On our Best Behavior". *Artificial Intelligence* [online], 212, pp. 27-35. <<https://doi.org/10.1016/j.artint.2014.03.007>> [Retrieved: 22 November 2024].
- MCCULLOCH, Warren S.; PITTS, Walter (1943). "A Logical Calculus of the Ideas Immanent in Nervous Activity". *The Bulletin of Mathematical Biophysics* [online], 5, pp. 115-133. <<https://doi.org/10.1007/BF02478259>> [Retrieved: 22 November 2024].
- SEARLE, John (1980). "Minds, Brains, and Programs". *Behavioral and Brain Sciences* [online], 3 (3), pp. 417-457. <<https://doi.org/10.1017/S0140525X00005756>> [Retrieved: 22 November 2024].
- TURING, Alan (1950). "Computing Machinery and Intelligence". *Mind* [online], 59 (236), pp. 433-460. <<https://doi.org/10.1093/mind/LIX.236.433>> [Retrieved: 22 November 2024].