Philosophy of artificial intelligence

Akash Deep
Biomedical Engineering
July 2021

The philosophy of artificial intelligence is a branch of the philosophy of technology that explores artificial intelligence and its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology, and free will. Furthermore, the technology is concerned with the creation of artificial animals or artificial people (or, at least, artificial creatures; see artificial life) so the discipline is of considerable interest to philosophers.

1. Can a machine display general intelligence?

The basic position of most AI researchers is summed up in this statement, which appeared in the proposal for the Dartmouth workshop of 1956 :- "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.“.

The first step to answering the question is to clearly define "intelligence".

Turing test :-

Alan Turing reduced the problem of defining intelligence to a simple question about conversation. He suggests that: if a machine can answer *any* question put to it, using the same words that an ordinary person would, then we may call that machine intelligent. A modern version of his experimental design would use an online chat room, where one of the participants is a real person and one of the participants is a computer program. The program passes the test if no one can tell which of the two participants is human. Turing notes that no one (except philosophers) ever asks the question "can people think?" He writes "instead of arguing continually over this point, it is usual to have a polite convention that everyone thinks". Turing's test extends this polite convention to machines:

•If a machine acts as intelligently as a human being, then it is as intelligent as a human being.

2. Arguments that a machine can display general intelligence

Dreyfus saw in the goals and methods of artificial intelligence a clear rationalist view on intelligence. Dreyfus, the most fundamental way of knowing is intuitive rather than rational. When getting expertise in a field, one is only bound to formalized rules when first learning the reasoning. After that, the intelligence is rather present as rules of thumb and intuitive decisions. The rational approach of AI is clear in the foundations of what is called *symbolic* AI. Intelligent processes are seen as a form of information processing, and the representation of this information is *symbolic*. Intelligence is thus more or less reduced to symbol manipulation. Dreyfus analyses this as a combination of three fundamental assumptions:

• the psychological assumption, which states that human intelligence is rule-based symbol manipulation.

• the epistemological assumption, stating that all knowledge is formalizable.

• the ontological assumption, which states that reality has a formalizable structure.

Gödelian anti-mechanist arguments:

Kurt Gödel proved with an incompleteness theorem that it is always possible to construct a "Gödel statement "that a given consistent formal system of logic could not prove. Gödel conjectured that the human mind can correctly eventually determine the truth or falsity of any well-grounded mathematical statement and that therefore the human mind's power is not reducible to a mechanism and is too powerful to be captured in a machine. However the modern consensus in the scientist and mathematical community is that actual human reasoning is inconsistent; that Gödel's theorems do not lead to any valid argument that humans have mathematical reasoning capabilities beyond what a machine could ever duplicate.

3. Can a machine have a mind, consciousness, and mental states?

This is a philosophical question, related to the problem of other minds and the hard problem of consciousness. The question revolves around a position defined by John Searle as "strong AI“. A physical symbol system can have a mind and mental states. The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds. All the above hypothesis are enough to prove that a machine can have a mind, consciousness and mental state but there are still several question that need’s to be answered such as can a computer program, running on a digital machine that shuffles the binary digits of zero and one, duplicate the ability of the neurons to create minds, with mental states (like understanding or perceiving), and ultimately, the experience of consciousness.

The computational theory of mind or "computationalism" claims that the relationship between mind and brain is similar (if not identical) to the relationship between a running program and a computer. The idea has philosophical roots in Hobbes (who claimed reasoning was "nothing more than reckoning"), Leibniz (who attempted to create a logical calculus of all human ideas), Hume (who thought perception could be reduced to "atomic impressions") and even Kant (who analyzed all experience as controlled by formal rules).

4. Other related questions

There are a few researchers who believe that consciousness is an essential element in intelligence.

The computational theory of mind or ”computationalism” claims that the relationship between mind and brain is similar (if not identical) to the relationship between a running program and a computer.

If the human brain is a kind of computer then computers can be both intelligent and conscious, answering both the practical and philosophical questions of AI. • In terms of the practical question of AI (”Can a machine display general intelligence? • In terms of the philosophical question of AI (”Can a machine have mind, mental states and consciousness?