

Philosophy of Artificial Intelligence

Pushkin Rathore 18111043

July 19th, 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. These factors contributed to the emergence of the philosophy of artificial intelligence. Some scholars argue that the AI community's dismissal of philosophy is detrimental.

- The philosophy of artificial intelligence attempts to answer such questions as follows:
 - Can a machine act intelligently? Can it solve any problem that a person would solve by thinking?
 - Are human intelligence and machine intelligence the same? Is the human brain essentially a computer?
 - Can a machine have a mind, mental states, and consciousness in the same sense that a human being can? Can it feel how things are?

Questions like these reflect the divergent interests of AI researchers, cognitive scientists and philosophers respectively. The scientific answers to these questions depend on the definition of "intelligence" and "consciousness" and exactly which "machines" are under discussion.

Important propositions in the philosophy of AI include some of the following:

- Turing's "polite convention": If a machine behaves as intelligently as a human being, then it is as intelligent as a human being.
- The Dartmouth proposal: "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it."
- Allen Newell and Herbert A. Simon's physical symbol system hypothesis: "A physical symbol system has the necessary and sufficient means of general intelligent action."
- John Searle's strong AI hypothesis: "The appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds."
- Hobbes' mechanism: "For 'reason' ... is nothing but 'reckoning,' that is adding and subtracting, of the consequences of general names agreed upon for the 'marking' and 'signifying' of our thoughts..."

Some scholars argue that the AI community's dismissal of philosophy is detrimental. In the Stanford Encyclopaedia of Philosophy, some philosophers argue that the role of philosophy in AI is underappreciated. Physicist David Deutsch argues that without an understanding of philosophy or its concepts, AI development would suffer from a lack of progress.

	Human–Based	Ideal Rationality
Reasoning–Based:	Systems that think like humans.	Systems that think rationally.
Behaviour–Based:	Systems that act like humans.	Systems that act rationally.

Four Possible Goals for AI According to AIMA

Main article: 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.[5] 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".[15] 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.

One criticism of the Turing test is that it only measures the "humanness" of the machine's behaviour, rather than the "intelligence" of the behavior. Since human behavior and intelligent behavior are not exactly the same thing, the test fails to measure intelligence. Stuart J. Russell and Peter Norvig write that "aeronautical engineering texts do not define the goal of their field as 'making machines that fly so exactly like pigeons that they can fool other pigeons'".