Philosophy of artificial intelligence

Sarita Kanwar

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.

- Are human intelligence and machine intelligence the same?

It only concerns the behavior of machines and ignores the issues of interest to psychologists, cognitive scientists and philosophers; it does not matter whether a machine is really thinking (as a person thinks) or is just acting like it is thinking.Arguments against the basic premise must show that building a working AI system is impossible because there is some practical limit to the abilities of computers or that there is some special quality of the human mind that is necessary for intelligent behavior and yet cannot be duplicated by a machine by the methods of current AI research.

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.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 behavior, 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.

Twenty-first century AI research defines intelligence in terms of intelligent agents.

- Intelligent Agent:An intelligent agent is a software program that can make decisions or provide services depending on its surroundings, human input, and previous experiences. These applications can be used to gather data on a regular, pre-programmed schedule or when the user prompts them in real time.

Artificial brain:

An MRI scan of a normal adult human brain Hubert Dreyfus describes this argument as claiming that "if the nervous system obeys the laws of physics and chemistry, which we have every reason to suppose it does, then we ought to be able to reproduce the behavior of the nervous system with some physical device".

Human thinking is symbol processing:

It implies both that human thinking is a kind of symbol manipulation (because a symbol system is necessary for intelligence) and that machines can be intelligent (because a symbol system is sufficient for intelligence).

Gödelian anti-mechanist aeguments:

Gödelian arguments In 1931, 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 (such as a high-level symbol manipulation program) could not prove.

More speculatively, Gödel conjectured that the human mind can correctly eventually determine the truth or falsity of any well-grounded mathematical statement (including any possible Gödel statement), and that therefore the human mind's power is not reducible to a mechanism.

This is provably impossible for a Turing machine(and, by an informal extension, any known type of mechanical computer) to do; therefore, the Gödelian concludes that human reasoning is too powerful to be captured in a machine.

Dreyfus: the primacy of implicit skills

Hubert Dreyfus argued that human intelligence and expertise depended primarily on implicit skill rather than explicit symbolic manipulation, and argued that these skills would never be captured in formal rules.

Dreyfus's argument had been anticipated by Turing in his 1950 paper Computing machinery and intelligence, where he had classified this as the "argument from the informality of behavior.

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?