Philosophy of Artificial Intelligence

Meda Guna Varma
18111036, Biomedical Engineering
National Institute of Technology Raipur.
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. 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 toAIMA

The arena of Artificial Intelligence constitutes of several aspects:

1. Intelligent Agent: An intelligent agent is a program that can make decisions or perform a service based on its environment, user input and experiences. These programs can be used to autonomously gather information on a regular, programmed schedule or when prompted by the user in real time.

2. Problem Solving: It is a part of artificial intelligence that encompasses a number of techniques such as a tree, heuristic algorithms to solve a problem. We can also say that a problem−solving agent is a result−driven agent and always focuses on satisfying the goals.

3. Knowledge: It is the information about a domain that can be used to solve problems in that domain. As part of designing a program to solve problems, we must define how the knowledge will be represented. A representation scheme is the form of the knowledge that is used in an agent.

4. Reasoning: The reasoning is the mental process of deriving logical conclusion and making predictions from available knowledge, facts, and beliefs.

5. Planning: It is about the decision−making tasks performed by the robots or computer programs to achieve a specific goal. The execution of planning is about choosing a sequence of actions with a high likelihood to complete the specific task.

6. Uncertain knowledge: When the available knowledge has multiple causes leading to multiple effects or incomplete knowledge of causality in the domain.

7. Machine Learning: It is a process that improves the knowledge of an AI program by making observations about its environment.

8. Communicating: It is a part of natural language processing, technologies such as machine translation of human languages, spoken dialogue systems like Siri, algorithms capable of producing publishable journalistic content, and social robots are all designed to communicate with users in a human−like way.'

9. Perception: in Artificial Intelligence it is the process of interpreting vision, sounds, smell, and touch. Perception is a process to interpret, acquire, select, and then organize the sensory information from the physical world to make actions like humans.

10. Acting: It refers to the action done by the algorithm in real world based on the processed data the algorithm has been fed.