PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

ISHIKA JAIN
5th Semester Biomedical Engineering Department
National Institute of Technology, Raipur

Guided by :
Prof. Saurabh Gupta

SUMMARY

The philosophy of artificial intelligence is a subset of the philosophy of technology that deals with artificial intelligence(in terms of creating an artificial creature) and its insights into **intelligence**,**ethics**, **consciousness**,**epistemology**, and **free will** and **logic**. The philosophy of artificial intelligence attempts to answer such questions as follows:

1 Can a machine display general intelligence?

According to the **Dartmouth's 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."
Turing Test interpretation of intelligence is as follows: If a machine acts as intelligently as a human being, then it is as intelligent as a human being.
Intelligent agent: "If an agent acts so as to maximize the expected value of a performance measure based on past experience and knowledge then it is intelligent."

1.1 Arguments in favour:

The brain can be simulated : According to **Hubert Dreyfus**,"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 : **Allen Newell and Herbert A. Simon** proposed that "A physical symbol system has the necessary and sufficient means of general intelligent action."

1.2 Arguments against the proposition:

Gödelian anti-mechanist arguments: Gödelian concludes that human reasoning is too powerful to be captured in a machine.
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.

2 Can a machine have a mind, consciousness, and mental states?

This is centered around the proposition of **John Searle** that-A **physical symbol system** can have a mind and mental states.A physical symbol system can act intelligently."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.

3 Is thinking a kind of computation?

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

Can a machine have emotions: If **"emotions"** are defined only in terms of their effect on behavior or on how they function inside an organism, then emotions can be viewed as a mechanism that an intelligent agent uses to maximize the utility of its actions.

Can a machine be **self-aware**: It is an essential human property that makes a character fully human. Turing argues in this regard that- "can a machine be the subject of its own thought?" Can it think about itself? Viewed in this way, a program can be written that can report on its own internal states, such as a debugger.

Can a machine be original or **creative**: Turing notes that, with enough storage capacity, a computer can behave in an astronomical number of different ways.It must be possible, even trivial, for a computer that can represent ideas to combine them in new ways.However humans will have the upper hand where artistic creativity is concerned.

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. With newer inventions and development in field of AI the gap between the human and AI are narrowing with this the importance of philosophical approach to AI development along with scientific approach is very important.