NATIONAL INSTITUTE OF TECHNOLOGY RAIPUR

Philosophy of Artificial Intelligence

Pankaj Kumar kindo, 19111038
5th Semester Biomedical Department, Artificial Intelligence

Guided by :
Prof. Saurabh Gupta

1 SUMMARY

The philosophy of AI deals with application of knowledge and its understanding of **intelligence**(described as the ability to perceive or infer information, and to retain it as knowledge), **ethics** (seeks to resolve questions of human morality by defining concepts such as good and evil, right and wrong, virtue and vice, justice and crime), **consciousness** (it is one's awareness of internal and external existence), **epistemology** (a branch of philosophy that deals with knowledge), **free will** (an ability to choose between various possible outcomes) and **logic** (analysis and appraisal of argument). Since the development of AI and its advancement deals with creation of artificial beings hence it is of great interest to philosophers.
There are certain question that are raised by some scholars regarding ignorance of philosophi- cal aspect of AI which may result in its dangerous and malicious usage. The answer for scholars lie on the **proposition** of AI(in philosophy a proposition is the non-linguistic bearer of truth or falsity) some of them are:

- Can a machine display general intelligence? this question has been answered by Alan Turning in **Turning Test** as if a machine acts as intelligently as a human being, then it is as intelligent as a human being. Latest AI research defines intelligence as an **agent** (an agent is someone or something which perceives and act in environment) with the definition of agent if an AI agent can maximize the expected value of a performance measure based on past experience and knowledge then it is intelligent.

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

- Is thinking a kind of computation? The **Computational Theory of Mind** states that the relationship between brain and mind is similar to computer and it running program. the Hobbes view "reasoning is nothing but reckoning" this means thinking is kind of arith- metic operation. In modern **computationalism** we call mental state as implementation of right computer program.

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.