

---

# Assignment 1 - Philosophy of Artificial Intelligence

Akshita Midha (akshitamidha08052001@gmail.com)

20 July 2021

---

## 1 INTRODUCTION

Philosophy of Artificial Intelligence is the subcategory of philosophy of technology which explores the idea of artificial intelligence and understanding its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology and free will. Philosophy of AI was a result of some important propositions proposed by different scientists in the early stage. Some of these important propositions were: 1. Turing's Polite Convention 2. The Dartmouth Proposal 3. Allen Newell and Herbert A. Simon's physical symbol system hypothesis 4. John Searle's Hypothesis 5. Hobbe's Mechanism Machine's capability of displaying general intelligence concerns the behaviour of machines and ignores the issue of interest to psychologist, cognitive scientists and philosophers.

## 2 INTELLIGENCE

**Turing's Test** Alan Turing managed to reduce the definition of intelligence as framed here wherein he suggested 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 in a manner that no one can identify that out of two people conversation one is a computer.' Thus, he brought up his idea for intelligence in simple words as 'If a machine acts as intelligently as a human being, then it is as intelligent as a human being.' One disadvantage of the Turing Test was that it only focused on the measuring of the 'humanness' of the machine's behavior, rather than the 'intelligence' of the behavior.

**Intelligent Agent** Intelligent Agent is something which perceives and acts in an environment and 'PERFORMANCE MEASURE' defines what counts as success for the agent. An agent is considered to be intelligent when it maximizes the expected value of a performance measure based on past experience and knowledge.

## 3 ARGUMENTS FOR THE DISPLAY OF GENERAL INTELLIGENCE BY MACHINES

**NERVOUS SYSTEM AND MACHINES** Hubert Dreyfus managed to bring in arguments against the ability of a machine to be able to express general intelligence as he compared to be able to express general intelligence as he compared the nervous system to that of the machine and stated 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 behaviours of the nervous system with some physical device.' Thus, scientists concluded that merely mimicking the functioning of a brain would in itself be an admission of ignorance regarding intelligence and the nature of the mind.

**SYMBOL PROCESSING** In 1963, Allen Newell and Herbert A. Simon proposed that 'Symbol Manipulation' was the essence of both human and machine intelligence. They stated it as: 'A physical system has the necessary and sufficient means of general intelligent action.' This claim was highly strong implying that 'Human thinking is a kind of symbol manipulation and machines can be intelligent.' Hubert Dreyfus described another version of this position and called it as The Psychological Assumption, wherein he stated that 'The mind can be viewed as a device operating on bits of information according to formal rules.' The arguments amongst scientists helped us reach a conclusion which says that human thinking does not consist solely of high level symbol manipulation and it doesn't show that AI is impossible, instead only that more than symbol processing is needed.

**ANTI-MECHANIST ARGUMENTS** In 1931, Kurt Godel proved using an incompleteness theorem that it is always possible to construct a Godel statement that a given consistent formal system of logic could not prove. Godel further placed his understanding that human mind can correctly eventually determine the truth or falsity of any well-grounded mathematical statement therefore the human mind's power is not reducible to mechanism. Godelian anti-mechanist arguments relied on the innocuous-seeming claim that a system of human mathematicians is both consistent and believes fully in its own consistency. Thus, Godelian concluded that human reasoning is too powerful to be captured in a machine.

**PRIMACY OF IMPLICIT SKILLS** Hubert Dreyfus argued that human intelligence and expertise depended primarily on implicit skills rather than explicit symbolic manipulation and argued that these skills would never be captured in formal rules. According to Dreyfus's ideas Computational Intelligence paradigms, such as neural nets, evolutionary algorithms and soon are mostly directed at simulated unconscious reasoning and learning. Historian and AI researcher Daniel Crevier appreciated the efforts made by Dreyfus as his mostly stated concepts were proved with passing time.

**STRONG AI AND WEAK AI** John Searle came forward with the concepts of Strong AI and Weak AI. He stated that 'A physical system can have a mind and mental states' for Strong AI and 'A physical symbol can act intelligently' for Weak AI. Igor Aleksander, Star Franklin, Ron Sun, Pentti Haikonen and few more researchers believe that consciousness is an essential element in intelligence.

**MIND AND CONSCIOUSNESS** Concept of mind and consciousness are used differently by different communities. According to new age thinkers "consciousness is an invisible, energetic fluid that permeate life and especially the mind which is similar to Bergson's *elan vital*. CONSCIOUS will be presented as a fully human character with intelligence, desires, will, insight, pride and so on. According to some researchers MIND CONSCIOUSNESS are used as a kind of secular synonym for the soul. According to Neurobiologists, the actual relationship between the machinery in our own heads and its collective properties such as the mind, experience and understanding.

**SEARLE'S CHINESE ROOM** John Searle conducted an experiment which was commonly known as 'SEARLE'S CHINESE ROOM' to further come to a conclusion to a point whether the COMPUTER can have a mind and mental states or not. The experiment has varied responses and they were as follows: 1. The system reply and the virtual mind reply 2. Speed, Power and Complexity Replies. 3. Robot Reply 4. Brain Simulator Reply 5. Other minds reply and the epiphenomena reply. Through this experiment he argued that consciousness can't be detected by examining the behavior of machine, human being or any other animal.

## 4 ANSWERS TO OTHER RELEVANT QUESTIONS

The computational theory of mind or "COMPUTATIONALISM" claimed that the relationship between mind and brain is similar to the relationship between a running program and computer. Emotions are basically an effect which is responsible for the kind behavior shown by the individual and it is observed that emotions are a mechanism that an intelligent agent uses to maximize the utility of its action and these are just devices used for channeling behaviors in a direction beneficial to the survival of one's species. Researchers who believed in the existence of soul argued that "Thinking is a function of man's immortal soul which was called as THE THEOLOGICAL OBJECTION by Alan Turing.

## 5 CONCLUSION

In the attempts to construct such machines we should not be irreverently usurping. Development of machines which will be intelligent is possible but bringing in them the characteristics of human beings bring in different statements and arguments from the scientists worldwide with a new theory to prove their point. Experiments, thesis and ideas together combined will bring in the best for the AI for the world.