

# Philosophy of Artificial intelligence

Submitted by -Alok Gautam

July 2021

## 1 Introduction

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.

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. 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

## 2 Turing test

Turing test, originally called the imitation game by Alan Turing in 1950,[2] is a test of a machine's ability to exhibit intelligent behaviour equivalent to, or indistinguishable from, that of a human. Turing proposed that a human evaluator would judge natural language conversations between a human and a machine designed to generate human-like responses. The evaluator would be aware that one of the two partners in conversation is a machine, and all participants would be separated from one another.

### 3 Can a machine display general intelligence?

Is it possible to create a machine that can solve all the problems humans solve using their intelligence? This question defines the scope of what machines could do in the future and guides the direction of AI research.

It only concerns the behavior of machines and ignores the issues of interest to psychologists, cognitive scientists and philosophers; to answer this question, it does not matter whether a machine is really thinking (as a person thinks) or is just acting like it is thinking.

"Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it .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 (or by the methods of current AI research . Arguments in favor of the basic premise must show that such a system is possible.

This is a philosophical question, related to the problem of other minds and the hard problem of consciousness. The question revolves around a position defined by John Searle as "strong AI": A physical symbol system can have a mind and mental states Searle distinguished this position from what he called "weak AI"

A physical symbol system can act intelligently. Searle introduced the terms to isolate strong AI from weak AI so he could focus on what he thought was the more interesting and debatable issue. He argued that even if we assume that we had a computer program that acted exactly like a human mind, there would still be a difficult philosophical question that needed to be answered

### 4 Consciousness, minds, mental states, meaning

The words "mind" and "consciousness" are used by different communities in different ways. Some new age thinkers, for example, use the word "consciousness" to describe something similar to Bergson's "élan vital": an invisible, energetic fluid that permeates life and especially the mind. Science fiction writers use the word to describe some essential property that makes us human: a machine or alien that is "conscious" will be presented as a fully human character, with intelligence, desires, will, insight, pride and so on.