

# Philosophy of Artificial Intelligence

Aman Khandwe, 19111006

July 19, 2021

## 1 Introduction

Artificial **intelligence** philosophy is a branch of technological philosophy that investigates artificial intelligence and its implications for **knowledge** and understanding of intelligence, **ethics**, **consciousness**, **epistemology**, and free will. Some of the most important propositions in AI philosophy are-Turing's "polite convention", The Dartmouth proposal, John Searle's strong AI hypothesis, Allen Newell and Herbert A. Simon's physical symbol system hypothesis, Hobbes' mechanism.

## 2 Can a machine display general intelligence?

This statement, which appeared in the proposal for the Dartmouth workshop in 1956, sums up the basic position of most AI researchers: "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it." To answer the question, the first step is to define "intelligence." According to Alan Turing, 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. According to **Intelligent agent** definition, 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.

### 2.1 Arguments that a machine can display general intelligence

1. The brain can be simulated-Hubert Dreyfus describes this argument by saying 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 behavior of the nervous system with some physical device".
2. Human thinking is a symbolic processing - In 1963, Allen Newell and Herbert A. Simon proposed that the essence of human as well as machine intelligence was "**symbolic manipulation**."
3. Dreyfus: the primacy of implicit skills-Hubert Dreyfus claimed that human intelligence and expertise were primarily based on implicit skill rather than explicit symbolic manipulation, and that these abilities could never be captured in formal rules.

### 3 Can a machine have a mind, consciousness, and mental states?

According to John Searle: **strong AI** is a physical symbol system that can have a mind and mental states and **weak AI** is a physical symbol system that can act intelligently. Searle's two points of view aren't particularly relevant to AI research because they don't directly address the question "Can a machine demonstrate general intelligence?" According to some philosophers, consciousness is an invisible, energetic fluid that pervades life, especially the mind. There are arguments stating that computers can't have a mind and mental states. For example- Searle's Chinese room, Leibniz' mill, Davis's telephone exchange, Block's Chinese nation and Blockhead.

### 4 Is thinking a kind of computation?

The computational theory of mind, also known as "**computationalism**," asserts that the relationship between mind and brain is similar to (if not identical to) that between a running program and a computer. Some variations of computationalism-

1. According to Hobbes-Reasoning is nothing but reckoning.
2. According to Stevan Harnad- Mental states are just implementations of the right computer programs.

### 5 Important keywords related to AI

1. Intelligence.
2. Knowledge.
3. Ethics.
4. Consciousness.
5. Epistemology.
6. Intelligent agent.
7. Symbolic manipulation.
8. Strong AI.
9. Weak AI.
10. Computationalism.