ASSIGNMENT 1

Apoorva Singh

July 2021

1 Philosophy of Artificial Intelligence

The philosophy of artificial intelligence is branch of philosophy of technology that deals with artificial intelligence and its future scopes.It deals with understanding of intelligence,ethics,consciousness,epistemology and free will.As this technology deals with creation of artificial animals or human beings;philosophical disciples are highly needed area to learn.

Some key questions that philosophy of artificial intelligence offers and their some discussions and arguments are:

- Can machines act intelligently?Can it solve any problem that a person would solve by thinking?

Alan Turing reduced the problem of defining intelligence to a simple question about conversation. He suggests 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. Twenty-first century AI research defines intelligence in terms of intelligent agents. An "agent" is something which perceives and acts in an environment. A "performance measure" defines what counts as success for the 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." Arguments that a machine can display general intelligence:

- The brain can be simulated
- Human thinking is symbol processing
- Are human intelligence and machine intelligence the same? Is the human brain essentially a computer?
- Can a machine have a mind, mental states, and consciousness in the same sense that a human being can? Can it feel how things are?

Consciousness, minds, mental states, meaning:For philosophers, neuroscientists and cognitive scientists, the words are used in a way that is both more precise and more mundane: they refer to the familiar, everyday experience of having a "thought in your head", like a perception, a dream, an intention or a

plan, and to the way we know something, or mean something or understand something[citation needed]. "It's not hard to give a commonsense definition of consciousness" observes philosopher John Searle.

Some important propositions in philosophy of AI are:

- 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..."