Philosophy of artificial intelligence

Name: Priyanshu Gupta
Roll no: 19111042
Branch: Biomedical Engineering

19 July, 2021

A two page summary on the article "Philosophy of Artificial Intelligence".

The exploration of artificial intelligence and its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology, and free will is termed as the **philosophy of artificial intelligence**. Factors such as the creation of artificial creatures and artificial life contributed to its emergence. The philosophy of artificial intelligence attempts to answer such questions as follows:

1 Can a machine display general intelligence?

Is it possible to create a machine that can solve all the problems humans solve using their intelligence? The basic position of most AI researchers is summed up in this statement by Dartmouth that "Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it.änd it does not matter whether a machine is really thinking or is just acting like it is thinking. Arguments do arise against whether the AI could be compared to a human mind.
The first step to answering the question is to clearly define ïntelligence".

Turing test for intelligence   Alan Turing reduced the definition of intelligence to a simple question that: if a machine and a person answer a question and the observer is unable to distinguish between them, then we may call that machine as intelligent as a human being.

Intelligent agent definition   An ägentïs something which perceives from the environment and acts accordingly. If it acts productively based on past experiences and data then it is intelligent. Here the disadvantage is that they can fail to differentiate between things that think and things that do not.

1.1 Arguments that a machine can display general intelligence

The brain can be simulated   One such model namely the halamocortical model that has the size of the human brain (10^{11} neurons) was performed in 2005 and it took 50 days to simulate 1 second of brain dynamics on a cluster of 27 processors.

Human thinking is symbol processing   First approach to AI is the symbol processing. It implies that human thinking is a kind of symbol manipulation and that machines can be intelligent because a symbol system is sufficient for intelligence. However **arguments against symbol processing** show that symbol manipulation is just not sufficient for artificial intelligence.

Gödelian anti-mechanist arguments   Kurt Gödel proved with an incompleteness theorem that it is always possible to construct a "Gödel statement"that a given consistent formal system of logic could not prove. Gödel conjectured that the human mind can correctly eventually determine the truth or falsity of any well-grounded mathematical statement and that therefore the human mind's power is not reducible to a mechanism and is too powerful to be captured in a machine.
However the modern consensus in the scientific and mathematical community is that actual human reasoning is inconsistent; that Gödel's theorems do not lead to any valid argument that humans have mathematical reasoning capabilities beyond what a machine could ever duplicate.

2 Can a machine have a mind, consciousness, and mental states?

A physical symbol system can act intelligently and at the same time can have a mind and mental states. This statement helped in distinguishing terms defined by John Searle as "strong AI" and "weak AI". The

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."

Consciousness, minds, mental states, meaning Every domain had their own definitions for these words. <u>Philosophers</u> call this the hard problem of consciousness. <u>Neuro-biologists</u> believe all these problems will be solved as we begin to identify the neural correlates of consciousness. Some of the harshest critics of AI do agree that brain is just a machine, and that consciousness and intelligence are the result of physical processes in the brain.

2.1   Arguments that a computer cannot have a mind and mental states

Searle's Chinese room Through his thought experiment Searle concludes that the Chinese room, or any other physical symbol system, cannot have a mind, mental states and consciousness require actual physical-chemical properties of actual human brains; in his words "brains cause minds".

Responses to the Chinese room Responses to the Chinese room emphasize several arguments such as the systems reply and the virtual mind reply, speed, power and complexity replies, robot reply, brain simulator reply and many more.

3   Is thinking a kind of computation?

According to **computational theory of mind** the relationship between mind and brain is similar to the relationship between a running program and a computer. In terms of the practical question of AI ("Can a machine display general intelligence?"), some versions of computationalism make the claim that Reasoning is nothing but reckoning. Also in terms of the philosophical question of AI ("Can a machine have mind, mental states and consciousness?"), most versions of computationalism claim that mental states are just implementations of (the right) computer programs.

4   Other related questions

4.1   Can a machine have emotions?

Emotions here can be viewed as a mechanism used to maximise the utility of actions.

4.2   Can a machine be self-aware?

More clearly speaking Can a machine think about itself? A program can be written such as a debugger however self-awareness is not limited to this.

4.3   Can a machine be original or creative?

Many proposals are there for a machine to act creatively in several manners while it seems likely that humans will have the upper hand where artistic creativity is concerned.

4.4   Can a machine be benevolent or hostile?

One issue with highly intelligent and completely autonomous machines is that machines may acquire the autonomy and intelligence required to be dangerous very quickly and it was also noted that some **computer viruses** can evade elimination and have achieved "cockroach intelligence."They also can be unpredictable if are given some degree of autonomous functions.

4.5   Can a machine imitate all human characteristics?

Turing listed many arguments of the form ä machine will never do X", where X can be having a sense of humour, tell right from wrong, make mistakes, etc but also believed that no such bounds can be set.

4.6   Can a machine have a soul?