
ASSIGNMENT-1

PHILOSOPHY OF ARTIFICIAL INTELLIGENCE

Name: BINAY KUMAR SETHI
Roll no: 19111021
Branch: Biomedical Engineering

20 July, 2021

1 Philosophy of Artificial Intelligence.

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.

2 Important propositions of AI

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

3 Can a machine display general intelligence?

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. Arguments in favor of the basic premise must show that such a system is possible.

Turing test for intelligence Alan Turing reduced the problem of defining intelligence to a simple question about conversation. The program passes the test if no one can tell which of the two participants is human. One criticism of the Turing test is that it only measures the "humanness" of the machine's behavior, rather than the intelligence of the behavior.

Intelligent agent definition An agent is something that takes in information from its surroundings and acts on it. It is intelligent if it acts constructively based on previous experiences and facts. The drawback is that they may be unable to distinguish between thinking and non-thinking items.

3.1 Arguments that a machine can display general intelligence

1) **The Brain can be stimulated-** This argument, first introduced as early as 1943. It is estimated that computer power will be sufficient for a complete brain simulation by the year 2029. A non-real-time simulation of a thalamocortical model that has the size of the human brain was performed in 2005 and it took 50 days to simulate 1 second of brain dynamics on a cluster of 27 processors.

2) **Human thinking is a symbolic processing-** In 1963, Allen Newell and Herbert A. Simon proposed that "symbol manipulation" was the essence of both human and machine intelligence.

3) **Gödelian anti-mechanist arguments** - Gödelian anti-mechanist arguments tend to rely on the

innocuous-seeming claim that a system of human mathematicians is both consistent and believes fully in its own consistency. After concluding that human reasoning is non-computable, Penrose went on to speculate that some kind of hypothetical non-computable processes involving the collapse of quantum mechanical states give humans a special advantage over existing computers.

4) Drefys-the primacy of implicit skills - The situated movement in robotics research attempts to capture our unconscious skills at perception and attention. Computational intelligence paradigms, such as neural nets, evolutionary algorithms and so on are mostly directed at simulated unconscious reasoning and learning. AI research has moved away from high level symbol manipulation, towards new models that are intended to capture more of our unconscious reasoning.

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

Searle introduced the terms to isolate strong AI (A physical symbol system can have a mind and mental states) from weak AI (A physical symbol system can act intelligently) so he could focus on what he thought was the more interesting and debatable issue. Neither of Searle's two positions are of great concern to AI research, since they do not directly answer the question -can a machine display general intelligence?

Consciousness, minds, mental states, meaning Every domain had their own definitions for these words. Philosophers call this the hard problem of consciousness. Neuro-biologists 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.

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

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

Responses to the Chinese room Several arguments are highlighted in the responses to the Chinese room, including the systems and virtual mind replies, speed, power, and complexity replies, robot reply, brain simulator reply, and many others.

5 Is thinking a kind of computation?

The relationship between mind and brain, according to computational theory of mind, is analogous to that between a running programme and a computer. Some versions of computational functions claim that Reasoning is nothing more than reckoning when it comes to the practical topic of AI ("Can a machine demonstrate general intelligence?"). In terms of the philosophical question of AI ("Can a machine have a mind, mental states, or consciousness?"), most computational functions assert that mental states are simply computer programme implementations.

6 Other related questions

1-Can a machine have emotions? Emotions in this scenario can be viewed as a mechanism used to maximise the utility of actions.

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.

3-Can a machine be original or creative? They can act creatively but still it will be less than that of a man.

4-Can a machine be benevolent or hostile? Machines can possess the intelligence required to be dangerous very quickly for instance some computer viruses can evade elimination and have achieved "cockroach intelligence." They also can be unpredictable if are given some degree of autonomous functions.

5-Can a machine imitate all human characteristics? Turing noted there are many things that a machine can't imitate to human like sense of humour, falling in love. But if they start possessing emotions it might be possible.

6-Can a machine have a soul? It depends on people believing and not believing on existence of soul. As it is a superstitious phenomenon one can't really conclude on it.