

Philosophy of Artificial Intelligence

1 Introduction

The philosophy of artificial intelligence explores artificial intelligence and its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology. There are many questions related to philosophy of AI which reflect the divergent interests of AI researchers, cognitive interests and philosophers. Questions mainly raised are: Can a machine act intelligently? Can it solve any problem that humans usually solve by their thought process? Can a machine feel, have consciousness, mental state like human beings?

1.1 Propositions

Turning's Polite Convention: If a machine behaves as intelligent as a human being, it is as intelligent as human being.

Dartmouth Proposal: It states that if an aspect or feature of intelligence can be so precisely described that a machine can be made to simulate it.

Allen Newell and Simon's Physical Symbol system: A physical system has the necessary and sufficient means of General Artificial Intelligence.

John Searle's strong AI hypothesis: appropriately programmed computer with the right inputs and outputs would thereby have a mind in exactly the same sense human beings have minds.

Hobbes's Mechanism: Reasoning is basically adding and subtracting, of the consequences of general names.

1.2 Questions and arguments related to AI

Can a machine have 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. For instance, machine learning, beginning with Turing's infamous child machine proposal essentially achieves the desired feature of intelligence without a precise design-time description as to how it would exactly work. The first step to answering the question is to clearly define intelligence.

Turing Test

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. Since human behavior and intelligent behavior are not exactly the same thing, the test fails to measure intelligence.

The Brain can be simulated

Most AI programs written between 1956 and 1990 used this kind of symbol. Modern AI, based on statistics and mathematical optimization, does not use the high-level symbol processing that Newell and Simon discussed.

Arguments against symbol processing

They do not show that artificial intelligence is impossible, only that more than symbol processing is required.

Summary

After concluding that human reasoning is non-computable, Penrose went on to controversially speculate that some kind of hypothetical non-computable processes involving the collapse of quantum mechanical states give humans a special advantage over existing computers. Existing quantum computers are only capable of reducing the complexity of Turing computable tasks and are still restricted to tasks within the scope of Turing machines. By Penrose and Lucas's arguments, existing quantum computers are not sufficient, so Penrose seeks for some other process involving new physics, for instance quantum gravity which might manifest new physics at the scale of the Planck mass via spontaneous quantum collapse of the wave function.

Searle distinguished this position from what he called weak AI

Human thinking is symbol processing Main article: Physical symbol system In 1963, Allen Newell and Herbert A. Simon proposed that "symbol manipulation" was the essence of both human and machine intelligence.

They wrote: "A physical symbol system has the necessary and sufficient means of general intelligent action.

" This claim is very strong: it implies both that human thinking is a kind of symbol manipulation (because a symbol system is necessary for intelligence) and that machines can be intelligent (because a symbol system is sufficient for intelligence). Another version of this position was described by philosopher Hubert Dreyfus, who called it "the psychological assumption": "The mind can be viewed as a device operating on bits of information according to formal rules.

"The "symbols" that Newell, Simon and Dreyfus discussed were word-like and high level — symbols that directly correspond with objects in the world, Most AI programs written between 1956 and 1990 used this kind of symbol.

Modern AI, based on statistics and mathematical optimization, does not use the high-level "symbol processing" that Newell and Simon discussed.

" A physical symbol system has the necessary and sufficient means of general intelligent action.

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. Neither of Searle's two positions are of great concern to AI research, since they do not directly answer the question

John Searle asks us to consider a thought experiment: suppose we have written a computer program that passes the Turing test and demonstrates general intelligent action.

can a machine display general intelligence?

. Before we can answer this question, we must be clear what we mean by minds, mental states and consciousness

The words mind and consciousness are used by different communities in different ways. For others, the words mind or consciousness are used as a kind of secular synonym for the soul. It's not hard to give a commonsense definition of consciousness observes philosopher John Searle. Some of the harshest critics of artificial intelligence agree that the brain is just a machine, and that consciousness and intelligence are the result of physical processes in the brain.

Can a machine be self-aware?

Self-awareness, as noted above, is sometimes used by science fiction writers as a name for the essential human property that makes a character fully human. Viewed in this way, a program can be written that can report on its own internal states, such as a debugger.

Can a machine have human characteristics?

Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humor, tell right from wrong, make mistakes, fall in love, enjoy strawberries and cream, make someone fall in love with it, learn from experience, use words properly, be the subject of its own thought, have as much diversity of behaviour as a man, do something really new. Turing argued about how these are disguised form of consciousness and naive consequences of it.

Can a machine be original or creative?

Turing reduces this to the question of whether a machine can "take us by surprise" and argues that this is obviously true, as any programmer can attest. He notes that, with enough storage capacity, a computer can behave in an astronomical number of different ways. It must be possible, even trivial, for a computer that can represent ideas to combine them in new ways. Douglas Lenat's Automated Mathematician, as one example, combined ideas to discover new mathematical truths.

Can a machine be hostile?

Main article: ethics of artificial intelligence This question like many others in the philosophy of artificial intelligence can be presented in two forms. "Hostility" can be defined in terms function or behavior, in which case "hostile" becomes synonymous with "dangerous". The latter is the question "can a machine have conscious states?"

such as intentions in another form. The question of whether highly intelligent and completely autonomous machines would be dangerous has been examined in detail by futurists (such as the Machine Intelligence Research Institute). Vernor Vinge has suggested that over just a few years, computers will suddenly become thousands or millions of times more intelligent than humans. They discussed the possibility and the extent to which computers and robots might be able to acquire any level of autonomy, and to what degree they could use such abilities to possibly pose any threat or hazard.

Can a machine have a soul?

Finally, those who believe in the existence of a soul may argue that "Thinking is a function of man's immortal soul." Alan Turing called this "the theological

objection”.

important keywords related to AI

- Intelligence
- Knowledge
- Agent
- Physical symbol system
- weak,broad,general AI
- Human mind
- Intelligent agents
- Syntax and sementics
- Consciousness
- hostile
- ethics
- Machine Learning
- rationality
- Computation
- Natural language processing
- environment
- logic
- facts
- Algorithm