# AI Philosophy

Philosophers have been around far longer than computers and have been trying to resolve some questions that relate to AI: How do minds work? Is it possible for machines to act intelligently in the way that people do, and if they did, would they have real, conscious minds? What are the ethical implications of intelligent machines?

## Weak AI VS Strong AI

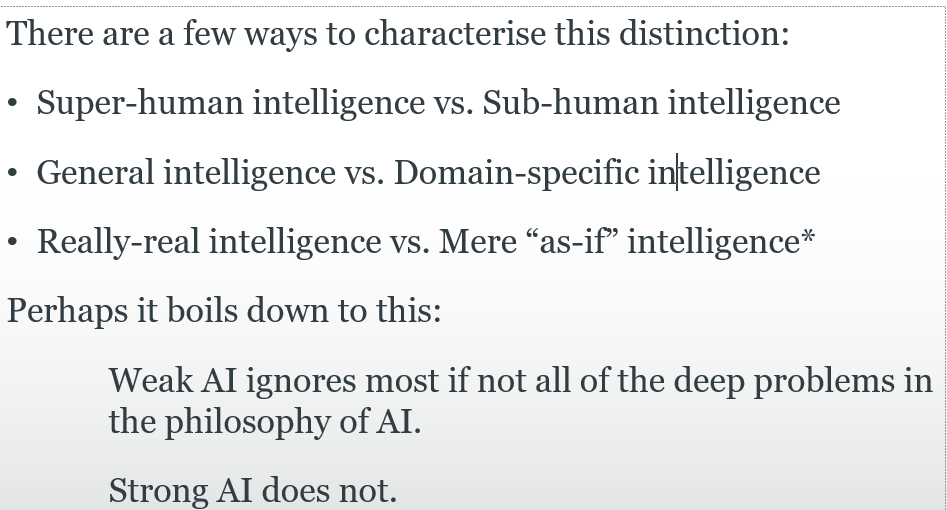
Weak and Strong AI are two different hypothesis about the degree to which machines can act intelligently as people do.

* Weak AI => the weak AI hypothesis is based on the fact that machines can act as intelligently as people but they do not actually think or are conscious. Therefore, they are just a mere simulation of intelligence (**simulating thinking, simulating brain**).
* Strong AI => the strong AI hypothesis is based on the fact that machines can act as intelligently as people do and they actually think and are conscious. Therefore, they have an actual brain.

Therefore, in a few words we can describe the debate Weak AI VS Strong AI with the following:

Simulation of intelligence VS Real intelligence

There are some other definitions of Weak AI and strong AI which are listed in what follows.



## Weak AI is the foundation of AI

AI was founded on the assumption that weak AI is possible as the proposal for the 1956 summer workshop that defined the field of Artificial Intelligence (McCarthy et al., 1955) made the assertion 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**”. Thus, AI was founded with the assumption that every aspect of learning or intelligence can be precisely described so that a machine can “read” the description and simulate intelligence.

## Can machine think?

Philosophers have posed the question whether machines can think long before the insurgence of computers.

According to the computer scientist Edsger Dijkstra the question of whether machines can think is about as relevant as the question of whether submarines can swim. The American Heritage Dictionary’s first definition of swim is “To move through water by means of the limbs, fins, or tail,” and most people agree that submarines, being limbless, cannot swim. The dictionary also defines fly as “To move through the air by means of wings or winglike parts,” and most people agree that airplanes, having winglike parts, can fly. However, neither the questions nor the answers have any relevance to the design or capabilities of airplanes and submarines; rather they are about the usage of words in English. (The fact that ships do swim in Russian only amplifies this point).

Now, we are going to see some objections to the Weak AI hypothesis.

## The argument from disability objection

The argument from disability states that an AI agent cannot do action X because it does not have a human brain. Naturally, action X is typically an action which is related to humans. Alan Turing listed the following actions as typically used as argument from disability against the weak AI hypothesis:

Be kind, resourceful, beautiful, friendly, have initiative, have a sense of humour, 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 man, do something really new.

However, such kind of argument is wrong as we know that many of the actions above are actually now performable by computers and in many cases even better than humans. Thus, it is wrong to assume that in order to perform some actions a human brain is needed. Furthermore, in the future, some actions which current AI agents do not excel at may become easy tasks for computers.

## The mathematical objection

The mathematical objection comes due to Godel Incompleteness theorem. Indeed, Philosophers such as J. R. Lucas (1961) have claimed that this theorem shows that machines are mentally inferior to humans, because machines are formal systems that are limited by the incompleteness theorem. However, mathematician Sir Roger Penros has written two books where he claims that such mathematical objections are false.

First, Godel’s incompleteness theorem applies only to formal systems that are powerful enough to do arithmetic. This includes Turing machines, and Lucas’s claim is in part based on the assertion that computers are Turing machines. This is a good approximation, but is not quite true. Turing machines are infinite, whereas computers are finite, and any computer can therefore be described as a (very large) system in propositional logic, which is not subject to Godel’s incompleteness theorem.

Secondly, humans have limitations as well. No human could compute the sum of a billion 10 digit numbers in his or her lifetime, but a computer could do it in seconds. Still, we do not see this as a fundamental limitation in the human’s ability to think. Humans were behaving intelligently for thousands of years before they invented mathematics, so it is unlikely that formal mathematical reasoning plays more than a peripheral role in what it means to be intelligent. Even humans cannot assert some sentences to be true while others can. Let’s suppose that the following sentence is true:

J. R. Lucas cannot consistently assert that this sentence is true

Given that this sentence is true then Lucan cannot assert the sentence otherwise the sentence would not be true. At the same time, other people can state that such a sentence is true. Therefore, this is an example of where a person cannot assert as sentence but others can. Even though this is true, we do not think that this limitation makes humans non intelligent.

Lastly, it is not possible to prove that humans are subject to Godel incompleteness theorem because currently there is no formalism that describe human brains.

I would like to add the following which is related to strong AI: the problem is not that there may be some problems that machines cannot solve or actions that cannot perform. The problem is that if a certain machine is really intelligent then it can solve all set of problems which have the same difficulty.

In other words, if a machine can solve a problem of difficulty n then if it is really intelligent it must be able to solve all problems whose difficulty is less than or equal to n.

## The argument from informality objection

The argument from informality objection is similar to the argument from disability as it claims that it is not possible to capture human’s reasoning through a set of rules. Since, machines can only follow rules then they cannot reason like humans.

This argument gave birth to the **qualification problem.**

The qualification problem in AI consists of the fact that not everything can be expressed in terms of logical statements. We have seen that this is true in the probability reasoning section. Indeed, the world is so complex that we do not know all factors that influence a certain outcome. Thus, we need probabilities in order to deal with the uncertain world. However, even though we do not know for sure, it may be that certain aspects of the world are not actually stochastic but rather deterministic but they look stochastic because it is a partial-observable environment for us.

The **qualification problem** affected the AI of that epoch which was based on symbols and logical reasoning on those symbols. The symbolic AI phase of that epoch is called “Good Old-Fashioned AI” (GOFAI). GOFAI’s assumption was that all intelligent behaviour can be captured by a system that reasons logically from a set of facts and rules describing the domain. In other words, GOFAI was based on logical inference, that is given a set of logical symbols comprising an initial knowledge (KB) find out whether other symbols can be inferred from the KB.

Therefore, these arguments from informality rather than claiming the lack of AI’s capabilities to simulate intelligence, they claim a lack of symbolic AI capabilities to represent intelligence.

An agent whose understanding of “dog” comes only from a limited set of logical sentences such as “Dog(x) ⇒ Mammal(x)” is at a disadvantage compared to an agent that has watched dogs run, has played fetch with them, and has been licked by one.

The opposite strategy to symbolic representation for developing AI agents is the **EMBODIED COGNITION.** The embodied cognition claims that it makes no sense to study only the brain separately but rather the whole body should be studied because it is through the body that an agent interacts with the environment and through such an interaction it is possible to learn. Therefore, within the EMBODIED COGNITION assumption when developing an agent, its sensors and its actuators become central in order to learn. Thus, AI should not focus on the agent function only but also on its architecture.

# Strong AI

There are many objections to the Strong AI hypothesis.

Many philosophers have claimed that a machine that passes the Turing Test would still not be actually thinking, but would be only a simulation of thinking. Again, the objection was foreseen by Turing. He cites a speech by Professor Geoffrey Jefferson (1949):

*Not until a machine could write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain—that is, not only write it but know that it had written it.*

Therefore, the main objections to the Strong AI hypothesis is that machines cannot have consciousness, intentionality or the capability of feeling emotions (phenomenology).

Alan Turing answer against such objections is the following:

He maintains that the question is just as ill-defined as asking, “Can machines think?” Besides, why should we insist on a higher standard for machines than we do for humans? After all, in ordinary life we never have any direct evidence about the internal mental states of other humans. Nevertheless, Turing says, “Instead of arguing continually over this point, it is usual to have the polite convention that everyone thinks.”

Thus, Alan Turing believes that the polite convention that we have in our daily lives about the fact that all people are able to think should be used with AI agents as well. Furthermore, as the time goes by and AI agents becomes increasingly intelligent the argument Weak AI VS Strong AI will dissolve away.

Indeed, a similar thing to what have happened to physics a long time ago may happen to AI in the near future:

A similar transition occurred in the years after 1848, when artificial urea was synthesized for the first time by Frederick Wohler. Prior to this event, organic and inorganic chemistry were essentially disjoint enterprises and many thought that no process could exist that would convert inorganic chemicals into organic material. Once the synthesis was accomplished, chemists agreed that artificial urea was urea, because it had all the right physical properties. Those who had posited an intrinsic property possessed by organic material that inorganic material could never have were faced with the impossibility of devising any test that could reveal the supposed deficiency of artificial urea.

## Mind-body problem

Assuming that a machine is intelligent when it acts as a human then the philosophical solution to the mind-body problem can clarify whether the Strong AI assumption is correct.

The mind-body problem is a debate concerning the relationship between the human mind and the body.

There are two main different answers to the mind-body solution and they are:

* Dualism => The dualism answer was devised by Renè Descartes and consists of the fact that the mind and the brain are distinct as the mind’s activity is a process which does not involve material properties. However, the problem that dualists face is how can the mind control the body given that they are separate?
* Monism (Physicalism) => Monism is based on the fact that the mind’s activity process is a physical process and as such consists of physical states. Thus, there is no distinction between the body and the mind. The problem that monists face is that they cannot figure out how to map the physical states of the brain to the mental states such as being in pain, enjoying a hamburger, knowing that one is riding a horse, or believing that Vienna is the capital of Austria.

The answer to the mind-problem is important for the Strong AI hypothesis because if Dualists are right then the Strong AI assumption is wrong. On the other hand, If Monism is correct then the Strong AI assumption is correct and all remains to figure out is how to create a brain through machines.

However, there is another claim about the mind-body problem which is Biological Naturalism which is an instance of Monism. This claim makes the possibility to recreate a real mind through digital machines an impossibility.

## Biological Naturalism (Searle)

According to which mental states are high-level emergent features that are caused by low-level physical processes in the neurons, and it is the (unspecified) properties of the neurons that matter. Thus, mental states cannot be duplicated just on the basis of some program having the same functional structure with the same input–output behaviour; we would require that the program be running on an architecture with the same causal power as neurons.

In other words, according to biological naturalism it is not possible for a machine to have a real mind even though it has the same input-output behaviour (given a certain input outputs something like a real mind would do) because mental states rely on some properties of the neurons. That is, according to Biological Naturalism, the mind is strictly related to the brain and so it is not possible to recreate a mind without a brain.

According to Searle, human brains may or may not be running something like an AI program, but if they are, that is not the reason they are minds. It takes more to make a mind—according to Searle, something equivalent to the causal powers of individual neurons. What these powers are is left unspecified.

## Qualia

Qualia is one of the aspects associated to having a consciousness. With qualia, it is meant one’s subjective experience. One’s subjective experience is a feeling to have a certain mental states (eating an hamburger, running, studying).

According to some philosophers it is impossible for humans to understand their own consciousness.

## Neat Vs Scruffy AI

“Neats vs. scruffies” is a sort of characterization used to describe two different artificial intelligence approaches. The “neats” prefer to advance in a way that is completely documentable and provable, in a method that is clear and logically supported. “Scruffies,” on the other hand, may embrace “fuzzier,” more diverse, or more ambiguous methods that support results. Therefore, “Scruffies” do not care about the formality of the AI model but only whether a particular model gives results.

Thus in conclusion Neat VS Scruffy AI can be summed up as:

**Provable AI methods VS Not-fully understood AI methods**