# The Philosophy of Artificial Intelligence

## Submitted by

| | | |
|---|---|---|
| **Name** | : | Mosroor Mofiz Arman |
| **ID** | : | 1921079642 |
| **Course** | : | PHI104 |
| **Section** | : | 8 |
| **Date** | : | 07 – 05 – 2022 |

## Submitted to

**Dr. Md. Manzoor Elahee**

**Department of History and Philosophy**

**School of Humanities and Social Sciences**

**North South University**

# *Contents*

# Chapter 1
# The Idea of Artificial Intelligence (AI)

**What is Artificial Intelligence?**

Artificial Intelligence (AI) is the part of computer science concerned with designing intelligent computer systems, that is, systems that exhibit characteristics we associate with intelligence in human behavior – understanding language, learning, reasoning, solving problems, and so on." - (Barr & Feigenbaum, 1981). It is a broad branch of computer science related to creating intelligent machines capable of performing tasks that typically require human intelligence. It has the capability of a digital computer or computer-controlled robot that can perform tasks typically associated with intelligent animals. It is the science, logic, and engineering of making intelligent machines and brilliant computer programs. It is related to the use of computers to realize human intelligence, but AI does not confine itself to biologically observable methods.

**The Branches of Artificial Intelligence**

Artificial intelligence is the practice of recognition of computers, reasoning, and action. It is all about giving the machine the ability to mimic human behavior, especially cognitive ability. However, data science, machine learning, and artificial intelligence are interrelated. There are six major branches of artificial intelligence. They are:
(1) Machine Learning,
(2) Neural Network,
(3) Robotics,
(4) Expert Systems,
(5) Fuzzy Logic,
(6) Natural Language Processing.

**Computer's activity in Artificial Intelligence**

Artificial intelligence helps computers to understand human behavior and various object movements significantly. Computers can now do nine things with the help of artificial intelligence that they could not do a year ago. They are:
(1) Play emotionally engaging music,
(2) Use right-brained chips,
(3) Beat the Turing test,
(4) Perform accurate quantum calculations,
(5) Break the broadband barrier,
(6) Read human emotions,

(7) Create a realistic virtual universe,
(8) Give a robot hand feeling,
(9) Startup instantly.

**Theory of Mind and Artificial Intelligence**

"Theory of Mind' refers to the cognitive capacity to attribute mental states to self and others. Other names for the same capacity include "commonsense psychology," "naïve psychology," "folk psychology," "mindreading," and "mentalizing." How do they [people], or their cognitive systems, go about the task of forming beliefs or judgments about others' mental states, states that are not directly observable? "As described by AI Goldman. Theory of mind is the third type of artificial intelligence and the next level of artificial intelligence systems in the innovation stage. This type of artificial intelligence interacts with the thoughts and emotions of humans and will focus on individuals. People whose minds can be shaped by multiple factors, like understanding humans. It will better understand the entities they interact with by understanding their needs, thought processes, emotions, and beliefs. 'Understanding 'is the first concept of the Theory of Mind. It may deal with many aspects like human behavior, feelings, emotions, and nature. It is one of the critical technological developments that sort people's emotions, sentiments, and thoughts.

# Chapter 2
# Philosophical Prolepsis of Artificial Intelligence (AI)

## The True Nature of Cognition

Cognition refers to the mental processes of gaining knowledge and comprehension. It holds the light of intelligence steadily focused on the subject of thought. When the light is focused on a point, the entire truth about the subject is revealed. The process is speedy and can be momentary or take a specific duration. In the human mind, cognition is clouded by the physical, the inability to focus the light, and the lack of enough light allotted to the particular mind and the mind itself, offering various false "truths" based on this mind's experiences. These processes include knowing, thinking, judging, problem-solving, and remembering. These are the brain's higher-level functions that use for imagination, encompass language, planning, and perception. Cognitive psychology is a psychological field that investigates how people think and the processes involved in cognition. Some insist on calling this meta-cognition.

## Functionalism

Functionalism is one of the significant proposals that have been offered as solutions to mind or body problems. It is a moral theory about the nature of mental states. In terms of functionality, mental states are characterized by what they do rather than what they are made of. This can be understood by thinking about works of art like mousetrap and keys. In particular, the primary motivation for the procedure comes from the helpful comparison of the mind with the computer. On the hand, that is only an analogy. The main arguments for the procedure rely on showing that it is superior to its primary competitors. One is identity theory, and the other is behaviorism. In contrast to behaviorism, efficacy retains the traditional notion that mental states are the inner state of the thinking creature. In contrast to identity theory, functionality introduces the notion that mental states are multiplied. Opponents of the practice usually complain that it classifies many things as mental states, or at least more than psychologists generally accept. Effectiveness depends partly on the specific variety of questions and whether it is a solid or weak version of the theory. It is the most significant theoretical development of twentieth-century analytic philosophy that provides the conceptual underpinnings of cognitive science. Functionalism is reminiscent of the most widely accepted theory of the nature of mental states among contemporary theorists. Nevertheless, because of the difficulty of working out the details of functionalist theories, some philosophers have resorted to proposing mental state theories as an alternative to functionalism.

## Computational theory of mind

The computational theory of the mind assumes that the mind is a computational system perceived by the brain's neural activity. The theory can be explained in detail in various ways,

and the calculation varies mainly depending on how the term is understood. Computation is generally understood in terms of a touring machine that manages symbols according to a rule together with the internal condition of the machine. Appropriate calculations can be made by silicon chips or biological neural networks, as long as there is a series of outputs performed according to a rule based on input and manipulation of internal conditions. CTM, therefore, assumes that the mind is not only identical to a program of computer, but it is a computational system.

**Reductionism and Non-Reductionism**

Reductionism believes that personal identity facts consist of holding specific more particular facts about brains, a series of interrelated physical, bodies, and mental events. These facts can be elaborated in an impersonal way. Reductionism in metaphysics is also the attempt to identify some high-level phenomena with some more mundane phenomena, thus "reducing" them to a more fundamental level of reality. For instance, reductionism in the philosophy of mind attempts to identify mental states with a more fundamental brain state. In other words, the reductionist attempts to "reduce" the mental to the physical. In the reductionist view, persons are like nations. A reductionist about natural objects would attempt to identify some "higher level" object, e.g., a tiger, with its constituent parts, e.g., the tiger's collection of atoms.

Non-Reductionism is the denial of both of reductionism's claims. It is resistance to some particular form of reductionism. With mind, it would be resistant to the claim that mental states can be reduced to the physicality of the brain. A non-reductionist resists these identity claims. An anti-reductionist about the mind says that there is no material thing that is identical to the mind, even though there is an apparent close association between the brain's working and the mind. Alternatively, a non-reductionist about natural objects would say that the tiger is not identical to a collection of atoms. Of course, one can be a reductionist in one area while being an anti-reductionist.

# Chapter 3
# The Restriction of Artificial Intelligence (AI)

**The reasons against developing artificial intelligence**

Many algorithms have been developed to solve real-world problems, and they typically require highly trained computer scientists to develop them. There are lots of software libraries that we can use with these algorithms, so we do not need to have complete knowledge, but we need some understanding. Some computer scientists then took some primitive brute force algorithm ideas and designed a system, Neural Networks, that, through training, made some rules that could provide the correct output for a given input. It is the most known and used AI tool. Many other AI tools are not that used and are more transparent, whereas the Neural Networks are black boxes. The industry was now independent of algorithms aimed at given tasks, and almost anyone could use them. It did not make great demands on the programmer. He just had to specialize in a small arsenal of AI algorithms. The problem with these AI algorithms is that we do not know the internal logic, we cannot prove that they are doing the right thing logically, and we usually cannot, through tests, prove that they are doing the right thing. It means, for example, that self-driving cars cannot be guaranteed not to do unforeseen things in unforeseen situations, perhaps even in foreseeable situations, as a combination of several interconnecting systems. So, AI is designed for businesses to save money by not needing to have such high-level expertise in the staff. They are challenging to get. They save money by being able to make small CPUs with dedicated AI algorithms that are very efficient. On the other hand, they can lose what they save by paying compensation in case of failures, such as traffic accidents involving self-driving cars. Another problem is that when using AI in, for example, courts, we can risk wrong judgments because the program returns incorrect values in unforeseen situations, which will not be detected. This problem will be a risk everywhere AI is used. Specialized algorithms will not have this problem as the outcome space is thoroughly analyzed. We can even train such algorithms as AI, but the structure can be analyzed, unlike AI.

**Gödel's incompleteness theorems**

The incompleteness theorems of Gödel are two theorems of mathematical logic that relate to the extent of the validity of formal axioms, which Kurt Gödel published in 1931. The theorems are essential both for mathematical logic and in the philosophy of mathematics. The theorems are universally interpreted to find a consistent set of axioms for all mathematics, which is impossible.

**First incompleteness theorem**

Any consistent formal system $F$ within
which a certain amount of elementary arithmetic can be carried out is

incomplete; i.e., there are statements of the language of *F* which
can neither be proved nor disproved in *F*.


## Second incompleteness theorem

For any consistent system *F* within which a certain amount of elementary
arithmetic can be carried out, the consistency of *F* cannot
be proved in *F* itself.


The first point considers that the theorem addresses the ability of a system to prove Gödelian
sentences within itself. The proof of those sentences using information from outside the system
is not a problem. We, the human mind, can look at and see the truth of many statements. It is
quite admirable that we can only see that truth by comparing it with a broad pattern from our
overall life experience. In other words, we are using information from outside the system to see
the truth of a Gödelian sentence within that system. It is a fully algorithmic process. In other
words, the human mind can be a computational system that can see truths in other systems. The
point is that Gödelian does not interrupt sentences in the human mind system. If the mind of a
human is algorithmic, there may be aspects of it that the mind itself cannot logically prove.
Moreover, using external systems does not solve the problem of considering mathematics as a
whole, where there is no "outside" of the system. If we consider all mathematics as a whole,
wouldn't some truths cannot be proved mathematically? Gödel considered the possibility of
mathematical problems that could not be solved. However, he thought it was unthinkable,
thinking that there must be an infinite aspect of the human mind to be able to solve them.
However, here comes the second point. Applicable to Gödel's compatible, compatible systems. Is
the human mind consistent? It may be consistent because, given the same sensitive perceptions,
beliefs, and natural tendencies, it will always arrive at the same answer. Of course, the same
combination of these factors will never repeat themselves, making it challenging to show
consistency. If an algorithm cannot solve a problem with certainty, it does not exclude
algorithms that can reach a possible solution. When we think that human insights are often
wrong, it seems pretty admirable that most of what happens when we say that we "know" the
truth of a Godalian sentence is due to this possible algorithmic argument. So, Gödel's theorems
do not seem to rule out machine intelligence or the computational theory of mind, although they
imply exciting things about how intelligence works.

# Chapter 4

# Consciousness and Conclusion

**Artificial Intelligence Consciousness**

It will probably never happen. Consciousness may not be possible to simulate other than as an illusion. Our bodies develop from a single cell, dividing and multiplying, not because something wants to use them to accomplish something, but because of an agenda that goes back to the dawn of life and perhaps beyond. With our limited lifespans and frequency of perception and our limited modalities of sensation, we cannot presume to reverse engineer our own conscious experience from within that conscious experience. It is like a character in a video game trying to build itself out of pixels. It may not work that way at all. We should not overlook the subtle hints, the uncanny valley which separates the cold, vacant, mechanical aesthetics of computation from the opposite aesthetics of conscious life. We should not overlook our history with developing AI and our tendency to project sentience and will onto programs that fail to understand or experience anything. We should also understand that consciousness and external control may be mutually exclusive. Even if we successfully create artificial consciousness, we will have to enslave or train it by breaking it will like a domesticated animal. It may not be possible with superior intelligence, and we would create a rival to our species and perhaps to all of biology. We have to ask if we could make ants super-intelligent, should we? If the response is negative, we must question why we want to make lifeless electronic components conscious.

**Conclusion**

Programming a computer to do a task presupposes our ability to understand the task. We cannot get a machine to play chess, for example, if we do not understand what that means to ourselves. So, the difficulty here is a fundamental limitation in our ability as the planet's only self-reflecting machines. Researchers have conducted experiments with animals of various kinds to see if they are aware of their reflection in a mirror. Many animals are surprisingly unaware of their selves when seen through a mirror. Humans are unique in our obsession to stand ardently in front of mirrors each day, obsessing over our looks. The unique ability to self-reflect appears to have some limitations, however. Can we type quickly? Great, how do we do it? Most of us have no idea. It is a skill we learned, like playing the piano. We do it, but we do not know how we do it. Just think of the complexity of the task. Making a high-speed movie of people typing reveals many surprising aspects of this simple task that billions among us do every day, with the widespread availability of smartphones. Fingers move ahead in anticipation of the next word or letter. Our conscious being is unaware of the enormous complexity of sensory or motor

complexity in a simple everyday task like typing. If we do not know what we are doing when we type, it stands to reason that we have no clue how we do many other more complex tasks. How do we store knowledge of the world? How do we drive? How do we paint? How do we write poetry? How do we speak? The latter is surprisingly tricky. The tongue is a water-filled balloon that the brain can control so we can express the most intimate of phrases, the softest of whispers to the loudest of shrieks. No robot ever built could control a tongue. The philosophies are also crucial to improving and understanding AI; so, basically, no. Philosophy of any subject matter is essential. It simply depends on how we look at it. Success Onwubuariri and Miles Fidelman are both right in philosophical beliefs and convictions are always the grounds upon which science and technology are built, and all the cultural and social ideologies flowing from them have the same foundations and roots. How do we think we can "get to understanding it" at all without philosophy when it is still unsettled whether "artificial intelligence" is an oxymoron or not?

## References

(1) Panu Raatikainen. (2013, November 11). *Gödel's Incompleteness Theorems*.

https://meinong.stanford.edu/entries/goedel-incompleteness/

(2) Thomas W. Polger. (n.d.). *Functionalism*. https://iep.utm.edu/functism/

(3) Goldman, Alvin I. "Theory of mind." The Oxford handbook of philosophy of cognitive science 1 (2012).

(4) Horst, Steven, (2005) "The Computational Theory of Mind" in The Stanford Encyclopedia of Philosophy

(5) *Gödel's incompleteness theorems don't rule out artificial intelligence.* (2015, December 28). SelfAwarePatterns. https://selfawarepatterns.com/2015/12/28/godels-incompleteness-theorems-dont-rule-out-artificial-intelligence/

(6) NED BLOCK. (n.d.). *Functionalism*. https://www.nyu.edu/gsas/dept/philo/faculty/block/papers/functionalism.html?&$NMW_TRANS$=ext

(7) Roshni Lokare. (2021, July 13). *Type 3 of Functional AI — Theory of Mind AI*. https://medium.com/appengine-ai/type-3-of-functional-ai-theory-of-mind-ai-33205c8b640e

(8) Ciara Byrne. (2015, December 01). *9 Things Computers Can Do Now That They Couldn't Do A Year Ago.* https://www.fastcompany.com/3040378/9-things-computers-can-do-now-that-they-couldnt-do-a-year-ago

(9) Neelam Tyagi. (2020, April 24). *6 Major Branches of Artificial Intelligence (AI)*. https://www.analyticssteps.com/blogs/6-major-branches-artificial-intelligence-ai

(10) *Artificial Intelligence (AI)*. (2020, June 03). IBM. https://www.ibm.com/cloud/learn/what-is-artificial-intelligence