International Journal of Social Science And Human Research

ISSN(print): 2644-0679, ISSN(online): 2644-0695

Volume 05 Issue 12 December 2022

DOI: 10.47191/ijsshr/v5-i12-31, Impact factor- 5.871

Page No: 5455-5457

An Exploration of Philosophical Ideas of Select Thinkers in Artificial Intelligence

Kumar Neeraj Sachdev[1], Aryan Mehra[2]

[1]Associate Professor of Philosophy, Department of Humanities and Social Sciences, Birla Institute of Technology and Science (BITS), Pilani Campus, Pilani – 333 031, Rajasthan, India

[2]Master of Science in Computer Science (MSCS), Computer Science Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States

ABSTRACT: We attempt to explore the presence and working of philosophical ideas of select thinkers in the whole architecture of artificial intelligence. This exploration delves into the philosophical connections relating to the proof of the existence of artificial intelligence. It further examines the philosophical perspectives of strong and weak versions of artificial intelligence, language processing and the inner working of artificial intelligence itself.

KEYWORDS: Artificial Intelligence (AI), Philosophical Ideas, Strong AI, Weak AI, Language Processing in AI

INTRODUCTION

Philosophy aims at understanding the bases of the object of inquiry. We aim to understand some philosophical thoughts of thinkers in the making and working of artificial intelligence. We show a few philosophical bases of the working ideas of artificial intelligence to enhance one's understanding of the working of it in the real world. We begin this endeavor with an overview of artificial intelligence, and we go on to demonstrate the relevance of select thinkers' ideas in the whole projection of artificial intelligence in the real world.

An Overview of Artificial Intelligence (AI)

We need to understand that artificial intelligence is not a new concept, nor is it complicated to understand. Alan Turing defined intelligence by answering another question which is famously called the imitation game (Alan Turing, 1950). Can a computer fool a human being and imitate as a male or a female human being? Many philosophers like Descartes also gave an early version of this thought, explaining how automatons or decision-making machines may imitate human decisions and actions, but may never do it as well as humans do.

John McCarthy, who is also known as the father of AI, is quoted as saying, "Intelligence is the computational part of the ability to achieve goals in the world. Varying kinds and degrees of intelligence occur in people, many animals, and some machines." He further defines AI as - "... the science and engineering of making intelligent machines, especially intelligent computer programs." (John McCarthy, 1981) Simply understood, AI is thus the ability of a computer program to mimic human decision-making and intellect and automate that process. There is simple math behind learning weights, numbers, or multipliers that provide the desired output when given the input. These "weights" or "multipliers" can be learned by calculus, or hard-coded by a programmer like static thresholds.

We see AI in our day-to-day endeavors and use AI everywhere without even thinking - spam filtering in emails, autocomplete and effective google searches, online shopping recommendations, playing chess against the computer, mobile banking, GPS recommended routes systems and so on and so forth. Right from unlocking your phone with face ID to asking using chatbots on helpdesks, we are surrounded by AI whether we realize it or not.

We may focus for a while on the role of media in misreporting AI and in turn demonizing it (Oscar Schwartz, 2018) to show a parallel but unrealistic picture of AI. Pop culture films like 'Terminator' are good fun and entertainment. But when it comes to imagining a future where AI will take over and destroy humanity is simply fiction. In fact, AI is not developed to the stage and probably ever will where such a coup is ever possible. A popular example is when Facebook engineers tested chatbots and assistants, it was noticed that the chatbots derived from English patterns that were too literal and started conversing in English words without syntax and semantics. While this was a technical barrier that engineers were trying to fix in language processing, editorials like the Sun were quoted as saying, "closely resembled the plot of The Terminator in which a robot becomes self-aware and starts waging war on humans." People started believing stories of AI systems creating their own language, far from reality because it was not even able to understand English fully at the time. Nonetheless, as we are surrounded by AI applications, it may be necessary to look from a philosophical bird's eye view on these things as they are developing ahead. This is why we analyze the importance and stance of some philosophical ideas in the making and working of artificial intelligence. We draw upon some select thinkers' ideas to show the variety of connections between philosophical ideas and the working of AI.

Existence of AI

We begin with Rene Descartes, a well-known contributor to modern western philosophy, algebra and analytical geometry, who puts forth the famous Latin dictum 'Cogito, ergo sum' (Rene Descartes, 1986), best translated as I think therefore I am. He thus gave a method to prove our existence through our ability to doubt, which is a form of thinking to establish on the grounds of intuitive certainty the fact that we exist.

Applying the same principle to AI, we see that AI algorithms can think and make decisions. Natural Language Processing is a field that has given rise to chatbots and conversational models that can doubt and prove their existence. Since doubt over one's existence is 'thinking', and to think is to possess 'being', we can infer that AI algorithms do exist. Still, their existence is conditional to the existence of human beings who have created them and have given them this ability to think. This is how a majority of rationalists would look at AI and debate its existence.

Strong AI and Weak AI

And in continuation, if we think about the consciousness of machines from a philosophical standpoint, we may consider John Searle's distinction between 'Strong AI' and 'Weak AI.' He regards weak AI as the branch of AI that is preprogrammed and can do only the tasks it has been specifically trained for (John Searle, 2007). An important example of these products includes Apple's Siri and other voice assistants, which can understand sentences because of pre-fed word meanings but cannot play around with its semantics and syntax beyond a certain limit. They simply use keywords like 'coffee' to extrapolate and search for coffee houses near you, for example but do not understand the significance of what the user is asking for. On the other hand, strong AI is closer to the human brain and is more complex. The actions of strong AI products are more unpredictable and learned, rather than being pre-programmed. A very interesting example of the same would be chess-playing algorithms. While we may predict or assume the computer's next move, there is no pre-programmed way of determining what the AI player may choose to do. It is as intelligent and as random as a human player's move.

Chinese Room Experiment

Searle further goes on to describe a 'Chinese room experiment' (David Cole, 2020). In this experiment, he describes a man who is given a set of instructions in English to converse and reply to someone in Chinese, although he himself does not know Chinese. In such a scenario, Searle asks, does anyone in the room know Chinese? All the entities, the man, the piece of paper, or the room, do not know Chinese. But for someone trying to converse with the room through the door, the room with all its members is intelligent enough to converse in Chinese. As explained above, he uses this experiment to strengthen his distinction between Strong AI and Weak AI. Knowing what to reply and do, and understanding the semantics of what was said, are very different things. This requires some more understanding of the working of language in AI, which we may explore with reference to language processing in AI.

Language Processing in AI

In regard to understanding the processing of language in AI, we may note that Wittgenstein, one of the greatest philosophers of the 20th century, worked on language and how humans process it. In his book "Tractatus Logico-Philosophicus", he famously stated that humans associate pictures or visualizations with thoughts or ideas ("making pictures of facts"), and that communication is simply the exchange of these pictures (Ludwig Wittgenstein, 1933). Misunderstandings, he said, occur when two people have a different picture of something or a deeper picture of something than what was intended to be said.

Using this philosophy to analyze the conversational aspect of AI, we notice that computers and natural language processing algorithms follow what we may call the "making vectors or numbers of facts". Many famous algorithms in AI, like the Word2Vec algorithm (Tomas Mikolov, 2013), use what we call the "bag of words" model. Here the order of words does not matter; what matters is only the word and its neighbours. Hence, every word's image' depends on its neighbours and usage. For example, 'love' and 'like' will have number representations closer to each other, because they occur in similar contexts. Hidden meanings, sarcasm, etc. are thus tougher for computers than it is for humans because the same words will always have the same vector representation. We may further appreciate this aspect of language in our focus on the inner working of AI.

Inner Working of AI

Freud, regarded as the pioneer of psychoanalysis, coined many terms that we use in our daily vocabulary. These terms include ego, superego, unconscious, subconscious, dream wish, etc. His works include "The Ego and the Id" (Sigmund Freud, 1961), among other groundbreaking works. He believed that all our thoughts and actions are balanced by three main aims or parts of our being - the ego, the superego, and the id. The id, as he explained, was the pleasure-seeking part of the brain, while the superego was the one that adheres to social norms and what is expected from us. The ego is supposed to comprehend the individuated reality and maintain the balance between the id and the superego.

Extrapolating this to a typical example of supervised learning, the id is the lessening of the loss function and obtaining great training accuracy. But merely doing that might seem mathematically favourable, but is not what is expected out of the training process. The result actually needs to be an image of the superego, and that is the validation accuracy. This is indeed the accuracy of the model on previously unseen data, making the model more robust to the outside world. The job of the machine learning engineer and the overall hyper parameter tuning process needs to be at the helm of both the training and validation accuracy. This is comparable to the job of the ego.

CONCLUSION

We may conclude that philosophical ideas are not that far from artificial intelligence or even computational linguistics as we could see in the presentation of connections between philosophical ideas and the working of artificial intelligence. We have made an attempt to show that the existence of artificial intelligence may be argued using Rene Descartes' philosophy, while it can be classified by John Searle's distinctions. Likewise, we see that Wittgenstein's ideology can draw parallels between how humans understand and process thoughts and how computers deal with language. Similarly, the overall apparatus of the Machine Learning system can be thought of as analogous to the "ego and id" concept of Sigmund Freud. Let us take the example of the chess-playing algorithm again. The fact that the algorithm can think about different moves on the board and prefer one move to another is proof of the fact that it exists (Descartes). It is classified as strong AI by Searle because it is as predictable or unpredictable as a human. It deals with the moves as probabilistic values and numbers, thus justifying thought as a number inspired by Wittgenstein's ideas. Lastly, the main aim of the algorithm is to win (superego). It is achieved by choosing the best move at every point to satisfy the immediate short-term goal (id). It is the engineer's job to ensure an adequate balance between the two (ego).

REFERENCES

1) Cole, David, "The Chinese Room Argument", The Stanford Encyclopedia of Philosophy (Winter 2020 Edition), Edward N. Zalta (ed.), URL:https://plato.stanford.edu/archives/win2020/entries/chinese-room/.
2) Descartes, René. (1986). Discourse on Method. New York: London: Macmillan; Collier Macmillan
3) Freud, Sigmund. (1961). The Ego and the Id. W W Norton & Co.
4) McCarthy, John. (1981). Some Philosophical Problems from the Standpoint of Artificial Intelligence. Readings in Artificial Intelligence
5) Mikolov, Tomas. et. al. (2013). Efficient Estimation of Word Representations in Vector Space, arxiv URL: https://arxiv.org/abs/1301.3781
6) Searle, John (November 12, 2007). What is AI? Basic questions. Stanford. URL: http://jmc.stanford.edu/artificial-intelligence/what-is-ai/index.html
7) Schwartz, Oscar, (2018). The discourse is unhinged: How the media gets AI alarmingly wrong URL: https://www.theguardian.com/technology/2018/jul/25/ai-artificial-intelligence-social-media-bots-wrong
8) Turing, Alan. (1950). Mind, Volume LIX, Issue 236, Pages 433–460
9) Wittgenstein, Ludwig (1933). Tractatus Logico-Philosophicus. [Reprinted, with a few corrections] New York: Harcourt, Brace