Proceedings of the 2nd International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/77/20240632

The prospect and metaphysical analysis of conscious artificial intelligence

Nian Lyu

University of Illinois Urbana-Champaign, Urbana-Champaign, 61801, United States

nianlyu2@illinois.edu

Abstract. Artificial intelligence, also known as AI, has led the trend of evolution in the past and future decades, and the potential of consciousness artificial intelligence emerges as a renovative field to address. The computer machine aims to process repetitive and tedious tasks for humans since its concept was first developed. Whether AI is conscious does not raise unprecedented discussion before the appearance of the concept of machine learning. After it appears, the machine, instead of merely passing in input and generating output, is able to learn while processing the information, which resembles a human's learning process. Therefore, this paper delves into the complex terrain of AI to explore the theoretical possibility of endowing machines with consciousness and addresses the future concerns and potentials of AI. Illustrating through the aspects of ethical concerns, metaphysical perspectives on consciousness, and the latest advancements in AI, the study critically examines whether machines can possess a consciousness similar to human understanding.

Keywords: Philosophy, Metaphysics, Artificial Intelligence, Consciousness, epistemology, Machine Consciousness.

1. Introduction

Chinese Room argument [1], published by Philosopher John Searle in 1980, suggests that it is impossible for AI, even strong AI, to have consciousness. Assuming that the computer in a room is a person who does not speak Chinese, and this person inside the room can answer every question the person outside asks by using the information provided which directly tells the person inside what to say, then the person outside would consider the person inside to understand Chinese; however, the person inside, which is a computer, never understands Chinese since he only does what he is told to do. This argument suggests that AI could never "understand".

Drawing from the Chinese Room argument, this paper illustrates initial skepticism surrounding the concept of conscious AI, suggesting that machines, an analogy to a person in a room following instructions to communicate in an unfamiliar language, lack genuine understanding despite appearing to comprehend. The advent of machine learning, however, challenges this notion by introducing a capacity for machines to 'learn' in a manner reminiscent of human cognitive processes, thereby reopening the debate on AI consciousness [2].

The paper further explores the philosophical underpinnings of consciousness through contrasting views: the materialistic perspective, which posits consciousness as a result of physical interactions within the brain, and David Chalmers's argument against materialism, emphasizing the need for a

DOI: 10.54254/2755-2721/77/20240632

fundamental theory of consciousness that transcends physicalist explanations. Additionally, the study references Cartesian Dualism to highlight the philosophical distinction between the mind and the body, suggesting complexity in replicating human consciousness in machines that go beyond physical mimicry.

Ethical considerations form a crucial part of the discourse, underscoring the urgency of addressing the moral implications of conscious AI development. Through a comprehensive review of literature spanning from the ethics of algorithmic intelligence to phenomenological and cognitive theories of consciousness, the paper aims to contribute to the ongoing dialogue on the future potentials and challenges of creating conscious AI. Ultimately, this analysis not only sheds light on the theoretical feasibility and ethical dilemmas of conscious AI but also calls for a nuanced understanding of consciousness that encompasses both material and metaphysical dimensions.

2. Theoretical background

Prior to endowing consciousness and mind to AI, it is significant to address the definition and metaphysical perspective of mind. The modern definition of the philosophy of mind is classified into two aspects.

2.1. Dualism

Cartesian Descartes, the first philosopher addresses the problem of interaction between our mind and body and promotes the mind-body theory or Dualism. In general, the definition of Dualism is: "In the philosophy of mind, dualism is the theory that the mental and the physical – or mind and body or mind and brain – are, in some sense, radically different kinds of things. Because common sense tells us that there are physical bodies, and because there is intellectual pressure towards producing a unified view of the world, … [3]" A well-known quote from Descartes implies his concept of mind-body theory, which is "I think, therefore I am". Descartes, being skeptical of everything in the physical world, concludes that the existence of consciousness is because a person cannot be skeptical of the skeptic itself, if he does, it would be a paradox. However, the critics of dualism puzzle Descartes and those who believe in the theory. The criticisms question if mind and body are independent of each other, how do mind and body interact such physical brain prompt to human's mind and thinking.

2.2. Physicist's view

The other is a perspective which considered to be materialistic and the mainstream in the current field of natural science called mind materialism. Specifically, the theory denotes that the process of consciousness is not mystic, instead, consciousness is just a chemical reaction between neurons in the human brain, and people's behaviors and feelings are current that the human body conveys to the brain. Furthermore, the theory manifests that creating conscious AI is possible in the future because the field of neuroscience and biology will make continuous progress until humans can define the human mind and create an AI to be able to think. Even though the process of how consciousness is constructed might be complex, ultimately, it is theoretically possible to manufacture such AI. Moreover, undertaking the process of continuously exploring the human creature, human beings resemble a super complex machine. Through Stanford Encyclopedia of Philosophy's description of how functionalism considers the relations between consciousness and the physical: "The functionalist often appeals to analogies with other inter-level relations, as between the biological and biochemical or the chemical and the atomic. In each case, properties or facts at one level are realized by complex interactions between items at an underlying level [4]." Functionalism [5], one of the theories of materialism, manifests that the conformation of consciousness derives from the function of thinking and surviving which portrays the overall thinking of materialism as consciousness is something that spontaneously, inherently, inevitably exists, which is compatible with individuals' being [6].

3. Critics of physicists' view

However, in The Conscious Mind: in Search of a Fundamental Theory by David Chalmers, who is an Australian philosopher and psychologist against the idea of materialistic perspectives of mind and

DOI: 10.54254/2755-2721/77/20240632

machines, he says: "Materialism is a beautiful and compelling view of the world, but to account for consciousness, we have to go beyond the resources it provides [7]". He suggests that artificial intelligence is not likely to self-generate consciousness, which is a fundamental theory. He illustrates that to define consciousness and create a conscious machine. Two problems need to be solved: the easy problem and the hard problem. The easy problem is the problem described above, which is the scientific aspects such as the computation power and hardware. Despite such artificial intelligence behaving the same as humans, such artificial intelligence might just be a philosophy zombie. The Stanford Encyclopedia of Philosophy defines zombie in philosophy: "Zombies in philosophy are imaginary creatures designed to illuminate problems about consciousness and its relation to the physical world… The most systematic use of the zombie idea against physicalism is by David Chalmers 1996 [8]" In other words, individuals cannot recognize whether an AI has consciousness, or it is just a philosophy zombie. Chalmers attempts to convey that a materialistic view of the mind does not demonstrate how consciousness could be endowed with artificial intelligence. The key aspect and justification he used is so-called the Hard Problem of Consciousness. The definition according to the Internet Encyclopedia of Philosophy: "The hard problem of consciousness is the problem of explaining why any physical state is conscious rather than nonconscious. It is the problem of explaining why there is "something it is like" for a subject in conscious experience, why conscious mental states "light up" and directly appear to the subject [9]." To simplify from another aspect, the hard problem of consciousness is a dilemma between feelings from the first-person point of view and the third-person point of view. For instance, the feelings of tasting something from the first-person cannot be portrayed from a third-person perspective, which makes an analogical comparison of phenomenal consciousness. Attempting to explain the subjective feeling of the phenomenon of consciousness is perplexing. There is an explanatory gap between the objective physical such as the interaction between the neurons, and atoms in our brain, and the subjective phenomenon of the mind like human's feelings, experiences, and mind which arouses the issue of how the physical yields human's subjective consciousness. Indeed, despite how developed neuroscience, biology, or science are, humans are unlikely to manufacture a machine with genuine understanding and consciousness. Instead, humans are more likely to create a pale philosophy zombie.

4. Ethical considerations

The field of AI ethics is the crucial foundation of future AI development. AI's ethical dilemma is considered a question that should be addressed immediately. AI ethics has two primary research fields. One is AI's ethical dilemma in current applications which primarily contributes to social issues, and another is the ontological issue of potential conscious machines.

4.1. Social issues

An instance of AI's ethical dilemma that arouses social issues is AI substituting blue-collar jobs. Two Brazilian economists analyze the phenomena of wage inequality: "The main hypothesis was that the demand and supply of skills had a central role in the growth of wage inequality, reducing the jobs and wage share of the median group. This occurred because such technologies complement occupations with high skills, replace medium-skilled laborers that perform routine tasks, and have little effect on low-skilled employees [10]". While the wages of the majority are affected, especially for individuals who perform routine tasks, the upper class remains uninfluenced. This phenomenon will result in extreme economic disparity between people. The society will merely consist of two classes, which are the upper class and the rest. Assuming this happens, individuals in the lower class find it impossible to step out of their original situation.

4.2. Ontological issues

The ontological problem is the most topical and concerning ethical issue of AI. Assuming such conscious artificial intelligence is successfully invented, then its social identity may seem groundless like scientific fiction, and a more practical issue is a dialectical and popular concept in movies and video games called "Cyborg", whose definition is: "a human being whose physiological functions are aided or

DOI: 10.54254/2755-2721/77/20240632

enhanced by artificial means such as biochemical or electronic modifications to the body [11]". The concept arouses ontological questions regarding human beings and machines. For example, if such human beings consist of mixed machines and human bodies (except his/her brain) and are allowed to neglect the pain, whether they will think like humans and whether human feelings will affect human thinking and consciousness. Even though the discussion of the ethical issues of genuine conscious machines seems fictitious, the discussion of humans with mechanisms on the physical body should be addressed since such a group will exist shortly. In order to give an answer, there is a need to address the fundamental question of whether the human brain associates with the human body in order to have consciousness, more specifically, whether the human body is merely a container that stores consciousness and conveys information to the human brain through the nerve system. A computer ceases to work if it lacks a motherboard; however, the human body is different from a computer. There are individuals who are able to survive without both legs and arms, and this creates an inductive hypothesis that the human eyes, kidney, stomach, and spine are substitutable as long as these organs perform the same function as their origin. In the future, keeping one's brain and connecting the brain with mechanics that function like eyes will still result in preserving one's thinking even though one might behave more rationally since they do not have feelings. According to this possibility, mental and consciousness seem to be the dominant part of nature and humans [12].

The crucial part of determining the potential of conscious machines is the relationship between the mental properties and the physical properties and conformation of the mind. According to David Chalmers' research, creating conscious machines is theoretically impossible because of the hard problem of consciousness. On the other hand, consciousness is something inherently spontaneous which suggests that generating a conscious machine is possible. Indeed, the ethical issue of AI is urgent and should be addressed immediately.

5. Conclusion

In conclusion, this paper critically engages with the frontier of artificial intelligence, probing the theoretical and ethical dimensions of equipping machines with consciousness. This exploration underscores the pivotal role of modern philosophy in shaping the development and understanding of AI. The insights drawn from metaphysical debates and philosophical discourse have profound implications for the evolution of artificial intelligence, offering a rich framework to navigate the complexities of consciousness and machine cognition. In the future, the prospect of AI continues to be both promising and fraught with ethical dilemmas. The integration of philosophical principles into AI research can guide people in addressing these ethical challenges, ensuring that advancements in AI are aligned with human values and societal well-being. By bridging the gap between philosophical inquiry and technological innovation, researchers can better anticipate the ramifications of conscious AI, fostering a future where AI ethics and human-centric approaches are at the forefront of technological progress. This paper highlights the necessity of a multidisciplinary approach, advocating for a future in which AI development is informed by ethical considerations, philosophical insights, and a commitment to beneficial outcomes for humanity.

References

[1] Ribeiro, A. (2021). Revisiting the Chinese Room: Looking for Agency in a World Packed with Archaeological Things. Cambridge Archaeological Journal, 31, 533 - 541.

[2] Simanowski, R., Brodmer, M. and Chase, J.G. (2019). On the Ethics of Algorithmic Intelligence. Social Research: An International Quarterly, 86, 423–447.

[3] Robinson, H. (2023). "Dualism", The Stanford Encyclopedia of Philosophy (Spring Edition), Edward N. Zalta & Uri Nodelman (eds.).

[4] Van Gulick, R. (2022). "Consciousness", The Stanford Encyclopedia of Philosophy (Winter Edition), Edward N. Zalta & Uri Nodelman (eds.).

[5] Wolfe, C.T. (2016). Materialism and 'the Soft Substance of the Brain: Diderot and Plasticity. British Journal for the History of Philosophy, 24(5), 963–82.

DOI: 10.54254/2755-2721/77/20240632

[6] Rosenthal, D. M. (1993). Thinking that one thinks. In M. Davies & G. W. Humphreys (Eds.), Consciousness: Psychological and philosophical essays (pp. 197–223). Blackwell Publishing.
[7] Chalmers, D.J. (1966). The Conscious Mind : in Search of a Fundamental Theory. New York :Oxford University Press.
[8] Kirk, R. (2023). "Zombies", The Stanford Encyclopedia of Philosophy (Fall Edition), Edward N. Zalta & Uri Nodelman (eds.).
[9] Weisberg, J. (2024). The Hard Problem of Consciousness. Internet Encyclopedia of Philosophy. Retrieved from https://iep.utm.edu/hard-problem-of-conciousness/.
[10] Acypreste, R.D., & Paraná, E. (2022). Artificial Intelligence and employment: a systematic review. Brazilian Journal of Political Economy.
[11] Heckathorne, C. (2023, September 20). cyborg. Encyclopedia Britannica. https://www.britannica.com/topic/cyborg.
[12] Lindia, M.S. (2023). Phenomenology of the Turing test: a Levinasian perspective. Journal of Communication.