

Comparing David Chalmer's and John Searle's Attitudes Toward Strong Artificial Intelligence

Kam Seng Li¹ Jiahe Tian² Beijun Xiong^{3, *}

¹Yali Lugu Middle School, Changsha, China

²Beijing Chenjinglun high school, Beijing, China

³Nanjing No.1 High School, Nanjing, China

* Corresponding Author Email: kamljc@ldy.edu.rs

Abstract. With the rapid development of artificial intelligence, the public and academic understanding of AI capabilities are also constantly changing. This change has triggered discussions about whether AI can have consciousness, understanding, or subjective experience. The definition of consciousness itself also remains controversial in philosophy and science, further complicating the issue. This article will discuss John Searle and David Chalmers' thoughts on whether artificial intelligence has consciousness from the three dimensions of the differences between philosophers on the development of machine learning, the relationship between consciousness and matter, and the nature of consciousness, and deeply analyze their thoughts on whether artificial intelligence has consciousness. Conscious views and opinions.

Keywords: Strong artificial intelligence, Chinese room thought experiment, biological naturalism, Machine learning, Consciousness.

1. Introduction

With the rapid development of artificial intelligence (AI), the question of whether strong artificial intelligence may have consciousness similar to humans has triggered widespread discussions. Especially with the technological advancement of machine learning and deep learning, AI has made significant breakthroughs in processing complex tasks. Many people have begun to think about whether future AI may have human-like consciousness. The discussion around this issue is not limited to the technical field, as philosophers have put forward different views on whether AI can be conscious. This article examines the contrasting views of two representative philosophers on this question, David Chalmers and John Searle.

Chalmers, as a functionalist, believed that consciousness is a manifestation of information processing [1]. He believes that with the continuous development of machine learning algorithms and complex neural networks, AI may master more advanced functions in the future, and may even produce a "super-intelligent AI" [1-4]. In his view, as long as the complexity and adaptive ability of the system reach a certain level, AI may have consciousness. This optimistic attitude is supported by many strong artificial intelligence advocates, who believe that the development of artificial intelligence may eventually break through current cognitive barriers and approach or even surpass human intelligence levels.

However, Searle was critical of this view. He refutes the idea that machines can be conscious through the Chinese Room thought experiment [2]. In Searle's view, AI only processes information according to predetermined rules. Like the people in the Chinese room, although it can output the correct answer, it does not understand the essence of the question. He advocates biological naturalism and believes that the emergence of consciousness relies on biological foundations rather than the operation of algorithms [3]. Therefore, it is impossible for AI, no matter how complex, to produce a consciousness similar to that of humans. Searle's criticism provides an important dimension of reflection on the philosophical foundations of artificial intelligence research, making the issue of consciousness about AI a cross-cutting topic in the fields of philosophy and technology [5-7].

Most of the existing research focuses on the discussion of the development of AI technology and the discussion of AI ethical issues, but there are relatively few studies that systematically compare different theorists' views on AI consciousness at the philosophical level. By comparing the views of Chalmers and Searle on the issue of strong artificial intelligence consciousness, this study not only helps to deeply understand the philosophical dimension of this issue, but also provides a clearer idea for the future development direction of artificial intelligence. The research of this article will fill the gaps in existing research through an in-depth analysis of the views of these two philosophers, and further explore whether AI may have consciousness, and what is the philosophical basis of AI consciousness.

This article uses a comparative writing method to analyze the different attitudes of Chalmers and Searle on the issue of strong artificial intelligence consciousness from three aspects: 1) Views on the potential development of machine learning; 2) Views on the relationship between consciousness and matter. Understanding; 3) Attitude towards the nature of consciousness.

2. Theory of Strong Artificial Intelligence

The theory of strong artificial intelligence believes that artificial intelligence can not only simulate or express human intelligent behavior, but also have real consciousness and understanding capabilities. According to this view, machines can not only perform tasks, but also have self-awareness, emotions and subjective experiences similar to humans. This theory holds that consciousness is achieved through computational processes, meaning that as long as the computational system is complex enough, machines may have the same mental state as humans.

2.1. Attitudes towards the Potential Development of Machine Learning Machines

Chalmers is positive and open to the development of machine learning, believing that as this technology continues to evolve, AI may have some form of consciousness in the future [4]. Complex neural networks and deep learning algorithms may allow AI to use deep neural networks to a certain extent to master some advanced (such as human-only) functions, such as image recognition, natural language processing, etc. [4]. In addition, AI's ability to self-reflect and adapt to the environment will also increase significantly, so that they may have a learning and adaptation process similar to humans. Chalmers also proposed that the "super-intelligent" AI that may appear in the future may be able to surpass humans in specific cognitive tasks (such as calculation, reasoning, etc.). This trend has already begun to appear: for example, as early as the last century, Top chess engines have been able to beat world champions in games. This may even pose a challenge to humans at the level of consciousness, raising the question that AI may have human consciousness.

Searle remains skeptical of the potential of machine learning. He proposed the difference between algorithms and consciousness: Although AI can improve algorithm and operating efficiency through machine learning, and adjust its own behavior, these operations are based on symbols and rules [2]. Therefore, this does not mean that they will gain true understanding or awareness of knowledge, because awareness often involves subjective experience, which cannot be simulated through mere computational processes. His Chinese room thought experiment effectively illustrates this point. He pointed out that AI, like the people in the Chinese room, only gives corresponding Chinese answers to the questions raised, but cannot truly understand the essence of the problem [2]. In addition, Searle proposed that consciousness needs a biological basis. This is explained in detail by his theory of biological naturalism [3]. Biological naturalism believes that consciousness is a biological function of the brain that is both physical and subjective. Although its origin and function can be explained through biology and neuroscience, the subjective experience itself cannot be completely reduced to a physical process. In this theoretical context, all features of consciousness (such as perception, emotion, and self-awareness) depend on the organism's nervous system, as well as complex biological processes such as neurons and neurotransmitters, which are other physical processes. Systems (such as AI) cannot simulate.

To sum up, Chalmers maintains a relatively optimistic attitude towards machine learning. His view supports that with the continuous optimization of machine learning technology, AI may further improve advanced functions such as image recognition and self-reflection, and may eventually evolve human consciousness. On the contrary, Searle believes from a biological naturalism perspective that AI cannot simulate consciousness because it lacks a corresponding biological basis (such as the brain and other neural organs). In addition, he pointed out the difference between algorithms and consciousness, that is, pure computational processes are not equivalent to subjective experience and understanding of things. In addition to his unique views on consciousness, Chalmers also has a different view of the relationship between matter and consciousness than John Searle, which leads to different views on strong artificial intelligence [1] [8-10].

2.2. The Relationship between Consciousness and Matter

In this regard, Chalmers, as a functionalist, gave the position that the state of mind is equal to the functional state, that is, consciousness is a derivative of matter [1]. However, unlike the identity theory, Chalmers does not think that only carbon-based brains like humans can produce consciousness. Functionalism believes that the material used to produce a kind of consciousness is not fixed, but optional, and it is not only the carbon-based human brain that can produce consciousness [1]. This view is called multiple realizability [1]. The human brain produces consciousness just like computer hardware drives software. Computers with different configurations run the same software, and the running programs are the same. For example, when running a WPS document on each computer, pressing the backspace key deletes characters instead of adding characters. If you want to control the body, you can generate consciousness through the human brain, or through a digital nervous system such as a brain-computer interface. Consciousness has such multiple realizability [1]. Although consciousness is derived from matter and has supervenient properties, it is not bound to a specific substance. In this way, consciousness is not necessarily bound to the human brain, and can be produced by different carriers. We believe that human consciousness can be transplanted into computers to achieve cyber immortality. Silicon-based hardware such as computers can also have consciousness [2]. This theory provides the basis for the theory of strong artificial intelligence. Since machines can think like humans and reach the level of human intelligence, robots can generate consciousness like humans and pass the Turing test, thus realizing the theory of strong artificial intelligence [2].

John Searle refuted it with the thought experiment Chinese House [2]. The logic of the refutation is that the state of mind is equal to the functional state [9]. As long as they are not equivalent, the refutation is successful. For this, only one counterexample is needed. The counterexample of functionalism in Chinese House is that there is a functional state but no state of mind. A person transmits Chinese from outside the closed house to inside. The house is like a computer, but the hardware inside is a human brain and an invincible dictionary [3]. The person in the house has reached the Chinese level of a native speaker through the dictionary and answers all Chinese questions. He can even say homophonic puns and perfectly produces a functional state. However, John Searle believes that although the person in the house has a perfect functional state, he does not have a mental state, that is, he has no understanding of the function he has produced [11]. For example, when outputting an apple, there is no imagination of apple in his mind, but apple directly corresponds to the word apple, so there is no mental state. When outputting Chinese jokes such as homophonic puns, I don't understand them either, so I don't laugh. So, computers that operate in the same mode as the Chinese Room have functional states but no mental states. They just match characters and then output them. Therefore, this thought experiment refutes the theory of strong artificial intelligence. The refutation believes that only substances such as the human brain can derive consciousness, and the emergence of consciousness is completely bound to the human brain [6]. Computers cannot produce consciousness. Searle used the example of water to illustrate this point. The wetness of water is produced by its material properties (the specific arrangement of water molecules), and consciousness is also a functional result caused by the biological material properties of the brain. Therefore, he

advocates that consciousness is a "characteristic of brain states" and is part of the material world. We think that it can also be refuted from the perspective of having mental states but no functional states. A newborn baby has no language or skills, but he has his own understanding of objects in his mind. When he sees adults running, he knows that legs can run, but he can't run and can't say that he knows that legs can be used for running. This is consciousness but no functional state.

David Chalmers believes that the mind-body problem is divided into simple problems and difficult problems, which are problems that can be solved by physics, chemistry, biology and other disciplines, and problems of subjective experience of consciousness [5]. The attribute dualism he advocates believes that mental states are equal to the spiritual attributes of material entities. The criticism of functionalism is that there is matter, such as stimulating a nerve, but there is not necessarily consciousness, such as feeling pain. At the same time, attribute dualism believes that consciousness is irreducible.

In general, Searle believes that consciousness is closely related to matter, but it is a manifestation of matter and cannot exist independently from the biological basis of the brain [6]. Chalmers believes that there is a profound relationship between matter and consciousness, but consciousness cannot be completely attributed to the product of material processes, but is an irreducible phenomenon. Therefore, he proposed that consciousness may be another basic existence parallel to matter [1].

2.3. The Nature of Consciousness

For Chalmers, consciousness is a foundation of the world, just like the laws of gravity and quantum mechanics. They are the most basic properties and laws. They are not used to explain more basic things. On the contrary, people explain and construct the world based on these laws. Chalmers believes that the same is true for consciousness. In his opinion, if the existing basic physical quantities (mass, time, space, charge, etc.) cannot explain consciousness, then consciousness should be something fundamental [1][5]. At the same time, he believes that consciousness is universal. The panpsychism he supports believes that everything, no matter how small, will have some primitive, subjective experience and feelings [5]. He believes that consciousness is the process of information processing [6]. As long as the system's processing method is complex enough and can reflect itself and adapt to the environment, it can be regarded as having consciousness. For example, he even believes that the characters in virtual reality may have a certain degree of consciousness if their reactions are complex enough.

On the contrary, in Searle's view, consciousness is more of an internal experience and understanding. Although AI can show intelligent behavior, it does not mean that they have consciousness or understanding [9]. Searle's Chinese Room Argument explains that AI does not really understand semantics, but only converts Chinese questions into Chinese answers. From the observer's perspective, AI seems to understand the problem, but in fact it does not [2][10].

Searle believes that information is not an objective entity, but is closely related to the observer's understanding and interpretation. He emphasized that the meaning and value of information often depend on specific contexts and backgrounds, rather than just a collection of data or symbols. There are two ways to understand the concept of information, one is observer-independent and the other is observer-dependent [3].

The intelligence of AI is observer-dependent. From the observer's perspective, its intelligence is only reflected in what kind of questions it answers and what actions it makes. There is nothing intrinsic. In the final analysis, it is to make them act like thinking through complex circuits. From the perspective of observer independence, it is not intelligent at all [7].

In general, the two have different understandings of the nature of consciousness. Chalmers believes that the ability to process information can reflect the complexity of consciousness, while Searle believes that machines are symbol operations and lack intrinsic understanding, and consciousness requires intrinsic understanding to support it.

3. Conclusion

Chalmers and Searle hold fundamentally different views on the nature of strong artificial intelligence (AI) and whether it can achieve consciousness. Chalmers adopts an optimistic stance, suggesting that with continuous advancements in machine learning, AI may eventually develop forms of human-like consciousness. He believes that consciousness could emerge from complex computational systems. In his view, as AI technologies evolve, they may acquire the ability to perform tasks that currently require human consciousness, such as self-reflection, decision-making, and understanding. Chalmers contends that consciousness is not necessarily tied to biology but might arise from the intricate organization of information-processing systems, like advanced neural networks.

On the other hand, Searle rejects this computational theory of consciousness and argues from a biological naturalism standpoint. He claims that consciousness is inherently tied to biological processes and cannot be replicated by computational systems alone. Searle's Chinese Room thought experiment illustrates his belief that symbolic manipulation, which AI relies on, does not equate to true understanding. In the experiment, a person in a room manipulates symbols to respond correctly in Chinese, but they do not actually understand Chinese. This, Searle argues, is analogous to how AI operates—processing inputs and outputs without genuine understanding. Therefore, he concludes that AI cannot achieve true consciousness, as it lacks the biological foundations, like neurons and brain chemistry, that underlie human consciousness.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

- [1] Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- [2] Searle, J. R. (1980). "Minds, Brains, and Programs." *The Behavioral and Brain Sciences*, 3(3), 417-424.
- [3] Searle, J. R. (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.
- [4] Chalmers, D. J. (2016). The singularity: A philosophical analysis. *Science fiction and philosophy: From time travel to superintelligence*, 171-224
- [5] Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3), 200-219
- [6] Chalmers, D. J. (2003). Consciousness and its place in nature. *The Blackwell guide to philosophy of mind*, 102-142
- [7] Searle, J. R. (1998). Consciousness, explanatory inversion, and cognitive science. In *Consciousness and Emotion in Cognitive Science* (pp. 139-196). Routledge.
- [8] Searle, J. R. (1990). The mystery of consciousness. *New York Review of Books*.
- [9] Searle, J. R. (1990). Is the brain a digital computer. In *Proceedings and Addresses of the American Philosophical Association* (Vol. 64).
- [10] Searle, J. R. (2002). *Consciousness and language*. Cambridge University Press.
- [11] Searle, J. (2014). *What your computer can't know*