computer sciences and mathematics forum

MDPI

Proceeding Paper

The Construction Approach of Machine Consciousness and Its Limitations †

Liang Wang and Ziyi Ma *

College of Marxism, Xi'an Jiaotong University, Xi'an 710049, China; text85@xjtu.edu.cn
* Correspondence: ma1214723351@xjtu.edu.cn
† Presented at Forum on Information Philosophy—The 6th International Conference of Philosophy of Information, IS4SI Summit 2023, Beijing, China, 14 August 2023.

Abstract: Along with the development of artificial-intelligence technology, the construction of machine consciousness has become a hot issue studied by scholars. Generally speaking, the research methods of machine consciousness are mainly divided into two approaches: the algorithm-construction strategy and the brain-like-construction strategy. The algorithm-construction strategy refers to the research of machine consciousness using mathematical and theoretical computations without considering the human-brain factor but by using the symbolic computation path. The brain-like-construction strategy refers to exploring the possibility of consciousness generation by machines through biophysical mechanisms on the basis of studying the mechanism of consciousness generation in human brain. In this paper, we will present these two paths of machine-consciousness construction and analyze their limitations.

Keywords: machine consciousness; construction approach; limitations

1. Introduction

Artificial consciousness (AC), also known as machine consciousness (MC), is a new field that has developed in recent years. It aims to understand consciousness by applying the principles and methods of information processing to machines, with the ultimate goal of giving machines human-like consciousness. From a technical point of view, machine consciousness refers to attempts by those who design and analyze informational machines to apply their methods to various ways of understanding consciousness and to examine the possible role of consciousness in informational machines [1]. From a philosophical point of view, machine consciousness is understood as an attempt to simulate and implement certain aspects of human cognition that are considered to be elusive and controversial phenomena of consciousness [2]. From the point of view of the research methods of machine consciousness, the research methods of machine consciousness are mainly divided into an algorithm-construction strategy and a brain-like-construction strategy. However, both types of construction strategies have certain limitations. First, the algorithm-construction strategy may have difficulties simulating the complexity and richness of human consciousness because it relies on mechanical imitation and simplification of cognitive processes. In addition, it is challenging to determine whether a machine exhibiting human-like consciousness actually possesses subjective consciousness or merely simulates it. Second, the brain-like-construction strategy faces certain obstacles due to our limited understanding of the complex workings of the brain. Despite advances in mapping the structure of the brain and identifying certain neural associations of consciousness, reproducing the complexity of the human brain, and achieving true simulations remains a daunting task. In the following, the author will specifically analyze these two machine-consciousness construction strategies and point out their limitations.

check for updates

Citation: Wang, L.; Ma, Z. The Construction Approach of Machine Consciousness and Its Limitations. *Comput. Sci. Math. Forum* **2023**, *8*, 16. https://doi.org/10.3390/cmsf2023008016

Academic Editors: Zhongzhi Shi and Wolfgang Hofkirchner

Published: 10 August 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

2. The Algorithm-Construction Strategy and Its Limitations

The algorithm-construction strategy simulates the cognitive process of humans by using the application of algorithms and models. This method generates the basis of conscious experience through simulation, with a view to the emergence of human-like consciousness in the machine. It mainly adopts the symbolic computation method, which has three operating mechanisms: one is reductionist rationalism, another is deductive logicalism, and the third is strong computationism [3]. Reductionism is a fundamental idea in the study of complex systems, which aims to gradually recurse a complex system into simpler subsystems until it reaches a point where each subunit can be fully described analytically. It is effective in dealing with some linear systems, but when faced with a nonlinear complex system such as the human brain, the overall behavior of the system cannot be fully understood only by analyzing the basic components of the system. Deductive logic is a type of formal logic. Its correctness depends on the accuracy of the deductive premise formula or axiom. But the existence of Godel's incompleteness theorem suggests that mechanical steps can never reach the wider perception of the mind. Strong computationalism is an expression of the extreme development of computationalism, divided into Turing's strong computationalism and philosophy-of-mind strong computationalism, which asserts that everything, including human consciousness, is calculable [4]. At present, the algorithmic-construction strategy also faces a series of developmental dilemmas: first, infinite database expansion eventually causes the database to "collapse". Although the current data-processing technology is becoming more and more advanced, in the field of distributed computing, data compression and filtering, as well as advanced database research and development, need to be further developed. Second, there has been difficulty in overcoming "semantic barriers". Human semantics is characterized by fuzziness and complexity. It is not only a simple combination of words and grammatical rules but also carries meaning levels, cultural nuances, and contextual dependencies. The language of artificial consciousness based on strong computationism is just machine language, and rather than understanding the semantics, it is a kind of imitation of human language. It cannot incorporate the whole system of language into its own thinking. Therefore, the problem of semantic barriers cannot be solved. Third, there is an inability to explain "common sense problems ". Human consciousness is a complex phenomenon involving various interrelated cognitive processes, including perceptual memory reasoning and consciousness itself (metacognition). When judging common-sense problems, it not only relies on reasoning of abstract concepts but also relies more on the summary of past experience. Algorithms are only good at pattern recognition and data analysis and lack practical experience in the real world. As a result, it has difficulty understanding abstract concepts or inferring implicit information in common-sense questions, focusing only on specific tasks. Therefore, it lacks the comprehensive and cognitive ability needed for consciousness creation.

3. Brain-like-Construction Strategies and Their Limitations

The brain-like-construction strategy proposes replicating the complex structures of the human brain in a machine, thereby replicating the corresponding functions of the human brain. By simulating the neural structures observed in the brain, researchers recreate human-like states of consciousness in a computational system. This approach relies heavily on advances in neuroscience and neuroengineering to uncover the complexity of the brain and translate it into machine-based models. There are two specific research approaches for the brain-like-construction strategy: one is the artificial-neural-network model, which focuses on building models by simulating the neuronal activity mechanism of the human brain; the other is the quantum-computing model, which uses the similarity between quanta and consciousness to build a bridge for the construction of relevant models.

3.1. Working Mechanism and Limitations of the Artificial-Neural-Network Model

The artificial-neural-network model was established based on the basic principles of biological neural networks. The basic principles of neural networks were used to

simulate the information-processing function of the brain in engineering practice, which is embodied in the construction of algorithms to solve practical problems, so that these algorithms would have interesting and effective computing power [5]. At present, artificial-neural-network models have developed a series of achievements, including (1) speech and audio recognition, (2) face and image recognition, (3) language- and information-retrieval processing, and (4) the creation of artistic products. Its operation conditions mainly include the following three: First, having a huge database as support. Second, the unprecedented development in the computing power of the computer. Third, the development of unsupervised learning models. However, the conditions under which the artificial-neural-network model can operate will also be the disadvantages that limit its development. First, relying on a huge database to run also means that the model has extremely high requirements for the capacity and content quality of the database, and the quantity and quality of data will directly affect the algorithm results of the model. Second is the lack of interpretability. Artificial neural networks are often referred to as black boxes because understanding and explaining their inner workings is extremely challenging. Artificial-neural-network models operate by learning complex nonlinear relationships between inputs and outputs, making it difficult to explain why such models make particular predictions. The lack of interpretability of artificial-neural-network models may pose ethical problems in areas where transparency and accountability are important.

3.2. The Working Mechanism of Quantum Computing Models and Its Limitations

Quantum theory of consciousness (QTOC) started from the discovery of several connections between consciousness and quantum mechanics in solving problems related to quantum measurements. For example, the highly integrated and highly differentiated nature of consciousness is similar to the nondeterministic nature of quantum correlations; the uncertainty of conscious activity is similar to the nonclassical uncertainty of quanta [6].

Although quantum computing is more computationally and descriptively powerful than classical computing, its uncertainty seems to be intrinsically linked to the mysterious nature of consciousness. We must also admit that the use of quantum-computing methods to achieve machine consciousness is unlikely at this stage. First, quantum mechanics has many unexplained problems within the framework of its own discipline, such as quantum gravity and the basic physical mechanism involved in quantum physics. Second, the reason that quantum mechanics can be linked to the study of consciousness is based on the premise that quantum states exist in the human brain, but there are still different views in academia on whether quantum computation can be performed in the human brain. Therefore, the way to achieve "machine consciousness" through quantum-consciousness theory requires waiting for further technological development and verification.

4. The Significance of Machine-Awareness Research

First, it is important to ease people's fear of machine consciousness. Through the above analysis of the current research strategy of machine consciousness, it can be seen that the current research on artificial intelligence still only involves the imitation of human knowledge and emotion. It is not yet able to solve the problem of intentionality in human consciousness, which is the "hard problem" to crack in terms of the human consciousness of self-awareness and feeling consciousness. As long as machines are unable to crack the problem of intentionality in human consciousness, the human fear of having a bionic man with human-like consciousness in the future will naturally dissolve. Furthermore, by delving into the study of consciousness, researchers can gain a deeper understanding of its various components and the mechanisms that underlie self-awareness in humans. This knowledge can inform the design and development of AI systems, allowing researchers to create machines that are more aligned with human intentions and values, reducing the likelihood of unintended consequences.

Second, the study of machine consciousness theory can achieve the effect of dispelling the mystery of human consciousness theory. By deciphering and translating the signals

sent by the human brain through algorithms, we can explore the operation mode of the cognitive mind, which has long existed as a black box in the human brain, thus demystifying cognition to a certain extent. For example, explicit models or theories of consciousness can be created by developing artificial systems that resemble human consciousness. These models can provide different frameworks and hypotheses for the underlying mechanisms of human consciousness. By comparing and contrasting the workings of artificial and human consciousness, researchers can gain new perspectives on the nature of subjective experience. Machine-consciousness research can also encourage phenomenological exploration of conscious experience, and researchers can examine the nature of perception and the connections between perception, attention, and consciousness by attempting to design artificial systems that exhibit not only intelligent behavior but also subjective experience.

Third, new theories and technologies in the study of machine consciousness can play a complementary role to traditional philosophical issues. For example, the discovery of brain–computer-interface technology can enrich and improve the theory of the "human-machine" subject relationship. For example, in medical ethics, the discovery of brain–computer-interface technology can enrich and improve the theory of human–machine agent relationships. At the same time, the discovery of brain–computer-interface technology is also promising for bridging the interpretative gap between the brain and the external environment. Machines can intervene in the brain and interpret the signals it sends out, thus translating the intention of the human mind. In terms of moral agency and responsibility, the study of machine consciousness has the potential to redefine the concept of moral agency in artificial-intelligence systems. By giving machines awareness, we might be able to assign a degree of responsibility for their actions, leading to ethical decision making and accountability. This advance is crucial to ensuring that AI systems behave ethically and can understand the consequences of their actions.

Author Contributions: Conceptualization, L.W.; methodology, L.W.; validation, L.W.; writing—original draft preparation, L.W. and Z.M.; writing—review and editing, L.W. and Z.M.; supervision, L.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by 2022 China Postdoctoral Science Foundation Project: Research on Ethical Risk Prevention of Intimate Intelligent Technology from the Perspective of Human-Robot Dependence Relationship (Project number: 2022M722553); 2023 Shaanxi Provincial Social Science Fund annual project: Research on the ethical risks and countermeasures of "asymmetric" intelligent human-robot interaction technology (Project number: 2023C032).

Institutional Review Board Statement: This study did not require ethical approval.

Informed Consent Statement: Not applicable.

Data Availability Statement: This study did not report any data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Google Encyclopedia. Machine Consciousness. Available online: http://www.scholarpedia.org/article/Machine_consciousness (accessed on 7 October 2022).
2. Wang, Y. The Possibility of Machine Consciousness. *J. Harbin Univ.* **2022**, *43*, 16–20.
3. Cheng, C. The working mechanism of artificial intelligence and its limitations. *Academics* **2021**, *1*, 189–196.
4. Cai, H.; Wang, K. The Humanistic Value in the Age of AI: Refutation of Strong Computationalism. *J. Humanit.* **2020**, *1*, 45–53.
5. Xiong, Z. *Introduction to Cognitive Science*; Huazhong Normal University Press: Wuhan, China, 2002; p. 208.
6. Chen, S.; Wan, X. Quantum Perspective: A New Approach to Consciousness Research. *J. Dialectics Nat.* **2014**, *36*, 1–5+125.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.