



Universidad Nacional de Catamarca - Facultad de Humanidades
Instituto de Investigación en Teorías del Arte y Estética
Revista Rigel – ISSN 2525-1945.

Artificial Intelligence and the Challenges of Consciousness: A Philosophical Approach in Light of John Searle's Thesis

Djamal Eddine Chikhaoui¹, Ahmed Kaabouche²

¹Lecturer (B), Nour El Bachir University Center – El Bayadh, Algeria

Email: d.chikhaoui@cu-elbayadh.dz

²Professor, Nour El Bachir University Center – El Bayadh, Algeria

Email: a.kaabouche@cu-elbayadh.dz

Received: 03/07/2026

Accepted: 14/09/2026

Published: 08/10/2026

Abstract

This article examines the philosophical problem raised by artificial intelligence with regard to its nature and limits through a critical analytical approach to the thesis of the American philosopher John Searle, who distinguishes between strong and weak artificial intelligence. The article seeks to test the hypothesis that artificial intelligence, despite its immense computational capabilities, remains incapable of attaining human consciousness and mental intentionality. The importance of this approach lies in the fact that it confronts us with fundamental philosophical questions concerning the nature of the human mind, the possibility of simulating it mechanically, and the limits of scientific knowledge in explaining mental phenomena. We adopted a critical analytical method in reading Searle's texts and comparing them with Turing's theses and those of strong AI, drawing on prominent foreign philosophical references such as the works of Hubert Dreyfus, Thomas Nagel, Ned Block, and Daniel Dennett. We conclude that Searle's distinction between simulation and consciousness constitutes a highly significant philosophical contribution. However, it does not negate the practical potential of weak AI in various fields; rather, it establishes epistemological and ethical limits to its uses and affirms the continuing centrality of the human role in knowledge and decision-making.

Keywords: Artificial Intelligence, Consciousness, Intentionality, Chinese Room Argument, Turing Test, John Searle, Philosophy.

1. Introduction



Artificial intelligence is no longer merely a technological development that changes the tools of production and communication; it has become an epistemic and cultural force that compels philosophy to reconsider fundamental questions: What is the mind? What is consciousness? To what extent can human thought be reduced to computational processes? Although this question is as old as the debate over the relationship between mind and body, it has acquired renewed urgency today because of machines' capacity to perform tasks once regarded as the exclusive domain of human beings. For this reason, Joudi Omran describes artificial intelligence as "both an urgent and legitimate matter" (Omran, 2024, p. 176), a description that requires the philosopher not merely to follow technological developments, but to question the theoretical foundations of this phenomenon.

The central problem here takes the form of a twofold question: Can a machine think as a human being thinks? Or is the intelligent behavior it produces no more than a formal simulation devoid of consciousness and intentionality? Responses have been divided between a position that views strong artificial intelligence (Strong AI) as a project capable of simulating or surpassing the human mind, and another that rejects this possibility on the grounds that human thought differs in nature from computational mechanisms. In essence, this dispute is not new; it extends the philosophical controversy concerning the relationship between mind and body and the possibility of reducing mental phenomena to material processes. Hubert Dreyfus was among the most prominent thinkers to draw attention to these limits, arguing that "unembodied machines cannot simulate higher mental functions" (Dreyfus, 1992, p. 3).

This article enters this debate from a specific angle: a critical analytical reading of the thesis of John Searle (1932–), one of the strongest opponents of the strong AI project. Through this reading, the article seeks to trace the concept of artificial intelligence and its historical development, examine Searle's distinction between strong and weak AI, discuss the Chinese Room Argument and the Turing Test, analyze the question of consciousness and mental states in their relation to machines, and identify the philosophical limits of artificial knowledge together with the ethical and existential dimensions arising from its development.

Accordingly, the problem from which the article proceeds may be formulated as follows: What is the nature of the philosophical limits of artificial intelligence? How can the relationship between human intelligence and artificial intelligence be understood in light of Searle's thesis? Can a machine achieve consciousness and intentionality, or is it limited to the formal simulation of intelligent behavior?

2. The Concept of Artificial Intelligence and Its Historical Development

Defining artificial intelligence is by no means a settled matter. The concept oscillates between an attempt to understand intelligent behavior and a project aimed at creating machines that simulate

such behavior. From this perspective, it may be defined as “the study of intelligent behavior, as well as an attempt to find the means by which intelligent behavior can be introduced into artificial machines” (Al-Ayadi & Ghanim, 2024, p. 4). Despite its plausibility, this definition remains general, since it raises a range of computational and cognitive questions: How can we produce a machine that behaves intelligently? What computational foundations are required for this? Does intelligent behavior necessarily imply understanding or consciousness? Thus, the practical objective of this science goes beyond building systems that are fast or accurate to the attempt to design systems that learn, understand, and adapt.

The theoretical roots of this field go back to the work of Alan Turing in the 1930s and 1940s. In 1936, Turing demonstrated that every computational process could, in principle, be carried out through a mathematical system, what became known as the universal Turing machine. This was followed by early attempts in the 1940s to design a system that “thinks.” Research subsequently expanded and accumulated, reaching a symbolic high point at the Dartmouth Conference in 1956, where John McCarthy formally coined the term “artificial intelligence.” Since then, ambitious projects have followed, including Japan’s Fifth Generation Computer Systems project in the 1980s, which sought to build computer systems with advanced reasoning capabilities that simulated the human mind.

At the analytical level, researchers have traditionally distinguished between two types of artificial intelligence: weak AI and strong AI. Weak AI remains confined to specific tasks, such as image recognition or natural language processing, whereas strong AI aspires to simulate the human mind in its entirety, including consciousness, perception, and intentionality. This distinction is not merely a technical classification; it is the core of the philosophical dispute addressed in the following sections.

In recent years, artificial intelligence has undergone striking developments with the emergence of large language models such as ChatGPT, Gemini, and Grok, which are capable of producing human-like texts and answering complex questions across multiple fields. This development has generated a new philosophical debate about the nature of these models and whether they possess some form of consciousness or understanding. David Chalmers argues that these models “may not be conscious, but they are not merely passive tools; rather, they are quasi-agents with quasi-genuine beliefs and desires” (Chalmers, 2025, p. 2). This position reveals the complexity of the relationship between artificial intelligence and consciousness, which is the relationship on which the analysis in this article focuses.

3. Strong and Weak Artificial Intelligence: Searle’s Distinction

John Searle is credited with formulating one of the most important philosophical distinctions in the field of artificial intelligence in his well-known 1980 article “Minds, Brains, and Programs,” in which he

directly criticized what was then called “strong artificial intelligence” (Searle, 1980, p. 417). Searle did not merely object to the ambition of this project; he reorganized the conceptual map of the entire field by distinguishing between two forms: weak AI, in which the computer is used as an effective tool for studying the mind, and strong AI, which assumes that a computer program is not merely a simulation of the mind but is a mind in the full sense, possessing intentionality and an understanding of the surrounding world.

The question Searle raises is not whether a machine can behave intelligently—something he does not deny—but whether it actually understands what it is doing. In this context, it is argued that “the computer or machine is merely a mechanical simulation of intelligent behavior and cannot amount to the genuine intelligence of the human being” (Al-Ayadi & Ghanim, 2024, p. 4). The difference between simulation and reality here becomes analogous to the difference between a simulated storm inside a computer and a real storm outside: the former does not make anyone wet, just as a simulation of understanding does not produce genuine understanding.

Searle’s rejection of strong AI rests on a central argument: computer programs manipulate symbols according to purely formal rules, and this type of processing cannot yield understanding or consciousness. In his terms, the program “has no cognitive states; it is merely a set of instructions executed by the machine without any awareness of what it is doing.” This opens the way to the Chinese Room Argument, one of the most famous philosophical arguments in this context, which will be addressed in the following section.

Searle’s objection is not confined to the technical dimension; it also extends to the broader philosophical position entailed by strong AI, whose proponents tend to claim that “the intelligible is the natural, and there is nothing outside nature” (Omran, 2024, p. 178). This position reduces existence to what can be computed and measured. Yet such reductionism encounters the limits of human knowledge itself, since there are domains of human experience—such as aesthetic, ethical, and religious values—that cannot be reduced to algorithms. This objection finds an echo in Thomas Nagel, who argued that “subjective experience cannot be reduced to physical processes” (Nagel, 1974, p. 435), implying that there are fundamental limits to any attempt to simulate the human mind mechanically.

4. The Chinese Room Argument and the Turing Test

The Chinese Room Argument occupies a central place in the philosophical debate on artificial intelligence. Searle introduced it in 1980 as his strongest objection to the claim of strong AI. The argument is based on a thought experiment that is simple in appearance but profound in implication: a person who does not know Chinese is locked in a room and provided with a set of

Chinese symbols and rules for manipulating them, enabling the person to produce appropriate responses to questions sent into the room in Chinese. From the outside, it appears that the person understands Chinese, whereas in fact the person understands nothing and is merely applying formal rules. Searle states in this context that “the observer may mistakenly believe that the person ‘understands’ Chinese when producing correct answers, whereas in fact he does not understand” (Searle, 1980, p. 418).

For Searle, the computer is in precisely the same position: it processes symbols according to programmed rules, but it does not understand their meanings. Genuine understanding requires consciousness and intentionality, two properties that cannot be realized in a purely computational system. Accordingly, a computer’s success in the Turing Test does not necessarily mean that it thinks or understands; it only means that it simulates human linguistic behavior.

The Turing Test, by contrast, is based on the idea that if a machine can answer certain questions in a way that is indistinguishable from human answers, it may be regarded as intelligent. Turing designed the test so that an interrogator communicates with both parties—the machine and the human—through written messages. If the interrogator cannot distinguish between them, the machine is considered to have passed the test. Searle, however, maintains that this test does not measure genuine understanding but only the capacity for external simulation, which is insufficient to regard the machine as thinking. Searle goes further by suggesting that “Turing’s design of the Turing Test incorporated a philosophical position within the philosophy of mind” (Al-Ayadi & Ghanim, 2024, p. 13), meaning that the test itself is not neutral but rests on prior philosophical assumptions about the nature of mind and intelligence.

The Chinese Room Argument has generated extensive debate in philosophical circles, dividing philosophers between supporters and critics. Among its most prominent critics is Daniel Dennett, who argues that consciousness is not a centralized phenomenon located in a single place in the brain, but rather the distributed product of multiple processes (Dennett, 1991, p. 101). Accordingly, searching for the “place” where understanding occurs in a computational system is misleading. Searle nevertheless insists that understanding requires self-consciousness, something he holds cannot exist in any computational system regardless of its complexity.

In recent years, new attempts have emerged to test the Chinese Room Argument in light of large language models. Some researchers argue that these models “do not possess internal causal models of the world” (Khamassi, 2024, p. 6), which means that they remain incapable of genuine understanding in Searle’s sense. Thus, more than four decades after the argument was first proposed, it continues to maintain a presence in contemporary debate, finding in each new technological development an opportunity for renewed examination.

5. Consciousness and Mental States Between Human Beings and Machines

Consciousness remains the most challenging property for any project that seeks to simulate the human mind mechanically. It is not merely a function that can be measured or tested from the outside, but a subjective experience lived from within by the subject. Searle describes it as “the mental states that we experience directly, such as pain, joy, desire, and belief,” states that cannot be reduced to computational processes. A computer can process information about pain, but it does not actually feel pain. This difference between processing and feeling is the difference between functional existence and conscious existence.

What further reinforces this distinction in Searle’s view is that mental states are characterized by intentionality, that is, by being directed toward a particular object or state of affairs. Thinking is always thinking about something, and believing is believing something. This kind of intentionality cannot exist in a computational system because the symbols processed by a computer do not refer to anything beyond themselves; rather, they acquire their meaning from the programmer who assigns it to them. Hence, Searle emphasizes that “consciousness is the most important property that distinguishes the human mind from other forms of artificial intelligence” (Al-Ayadi & Ghanim, 2024, p. 13), which leads him to question the possibility that a machine could reach the level of human consciousness.

In his famous article “What Is It Like to Be a Bat?”, Thomas Nagel posed a fundamental question about the nature of subjective experience: Is there “something it is like” to be a particular kind of being? Nagel argues that subjective experience “cannot be reduced to physical processes” (Nagel, 1974, p. 437). This means that even if we were able to simulate a bat’s behavior with complete precision, we would never know what it is like to be a bat. The same applies to a machine: it may simulate human behavior, but it will never experience what it is like to be human.

Ned Block adds another dimension to this debate by arguing that “brain waves generated by electrochemical processing may be necessary for consciousness” (Block, 2025, p. 298), suggesting that consciousness may be intrinsically linked to a biological substrate rather than being merely a computational function that can be implemented in any system. This position converges with Searle’s rejection of strong AI while adding a more specifically biological dimension. Block also notes that “artificial intelligence and simple animals are competitors when it comes to the attribution of consciousness” (Block, 2025, p. 299), opening the way for thought-provoking comparisons between artificial systems and simple biological organisms.

6. The Philosophical Limits of Artificial Knowledge

Artificial knowledge raises a number of philosophical problems concerning its nature and limits. On the one hand, artificial intelligence may be said to “aim at understanding the nature of human intelligence and simulating it through programs operating with specific algorithms on a computer” (Al-Ayadi & Ghanim, 2024, p. 13). On the other hand, the question remains whether such simulation is capable of reaching the essence of human knowledge. Hubert Dreyfus points out that “human knowledge is partly tacit, and therefore cannot be fully expressed and incorporated into a computer program” (Dreyfus, 1992, p. 15). This means that there are fundamental limits to artificial knowledge that concern not computational capacity, but the very nature of human knowledge.

Joudi Omran notes that artificial intelligence “inclines toward the position that the intelligible is the natural, and that there is nothing outside nature” (Omran, 2024, p. 178), a philosophical position that reduces existence to what can be computed and measured. Yet such reductionism encounters the limits of human knowledge itself, since there are domains of human experience—such as aesthetic, ethical, and religious values—that cannot be reduced to algorithms. Hence the need to establish an epistemology of artificial intelligence that takes these limits into account, neither exaggerating its promises nor ignoring its possibilities. The knowledge produced by artificial intelligence is probabilistic in essence, based on statistical patterns extracted from training data, and is vulnerable to bias and error.

Zouhir Ben Ali points out that “artificial intelligence offers promising opportunities in historical research for collecting scholarly material, analyzing texts and sources, and translation, while it suffers from limitations in accessing traditional (paper-based) sources in particular” (Ben Ali, 2025, p. 216). This limitation is not confined to historical research; it extends to various fields of human knowledge that require contextual understanding and qualitative interpretation that cannot be reduced to data processing. John Haugeland emphasizes that artificial intelligence “is based on a very good idea, which may or may not be true” (Haugeland, 1985, p. 3), namely the idea that human thinking and machine computation are “fundamentally identical.” Yet Haugeland himself questions this equivalence and argues that the human mind does not operate like traditional artificial intelligence, but rather resembles complex neural networks (Haugeland, 1985, p. 12).

7. The Ethical and Existential Dimensions of Artificial Intelligence

The problem of artificial intelligence is not limited to epistemological and philosophical aspects; it also extends to highly significant ethical and existential dimensions. From an ethical standpoint, AI applications raise questions about civil liability for the harm these applications may cause. Saif bin Nasser Al-Maamari notes that “civil liability for artificial intelligence applications remains a matter of debate and controversy among legal scholars in the absence of legal regulation clarifying the basis of civil liability for such applications” (Al-Maamari, 2024, p. 38). This legislative gap creates major

legal and ethical challenges, especially as reliance on these applications increases in sensitive fields such as medicine, transportation, and finance.

From an existential perspective, artificial intelligence raises concerns about the future of human beings and their role in a world where the automation of tasks is increasing. Joudi Omran states that “machines will surpass human mental capacities in the future, and many indicators suggest that machines will develop themselves autonomously” (Omran, 2024, p. 176), raising questions about the fate of human beings amid these transformations. Will we witness the disappearance of certain human jobs and skills? Or will artificial intelligence remain merely a tool in the service of humanity? Daniel Dennett suggests that “the best reason for believing that robots may one day become conscious is that we humans are a kind of machine, and we are conscious” (Dennett, 1991, p. 431). Yet this position overlooks the fundamental differences between a machine and a living organism, particularly with regard to consciousness and intentionality.

In this context, another ethical question emerges concerning bias in artificial intelligence systems. Numerous studies have shown that such systems may reproduce biases present in training data, leading to unfair outcomes for certain groups. Nazim Hassan Rashid notes that “bias is an inherent human trait and can be reflected in everything we create, especially when it comes to artificial intelligence” (Rashid, 2023, p. 438). This makes it necessary to establish an ethics of artificial intelligence that ensures fairness, transparency, and accountability. Rashid further notes that “numerous studies have emphasized the importance of auditing bias in artificial intelligence in order to reduce the risks associated with potential algorithmic bias” (Rashid, 2023, p. 430). This means that the ethical question is not confined to theoretical aspects but extends to practical practices that require oversight and auditing mechanisms to ensure the ethical and responsible use of artificial intelligence.

8. Conclusion

In conclusion, it can be said that the philosophical problem of artificial intelligence remains open to profound questions concerning the nature of the human mind, the limits of knowledge, and the possibility of mechanically simulating consciousness. Through our approach to John Searle’s thesis, we have concluded that the distinction between strong and weak artificial intelligence remains a fundamental philosophical distinction that helps us understand the epistemological and ethical limits of this technology. Simulation is not consciousness, and intelligent behavior is not necessarily evidence of understanding.

In light of this, weak artificial intelligence may be regarded as a valuable tool in the service of human beings across multiple fields, whereas strong artificial intelligence remains a controversial

philosophical project that cannot be accepted unless the possibility of consciousness arising in a mechanical system is demonstrated, something that has not yet been established. This means that the human role in knowledge and decision-making remains central and indispensable.

We have also concluded that the ethical and existential dimensions of artificial intelligence require the establishment of a new ethics that takes the distinctive nature of this technology into account, sets controls on its uses, and ensures that it does not become biased or deviate from human values. This makes it necessary to strengthen philosophical research in this field and encourage dialogue among philosophers, scientists, engineers, and policymakers in order to ensure the responsible development and use of artificial intelligence in ways that serve humanity rather than harm it.

9. References

Arabic References

1. Al-Ayadi, Fatiha, & Ghanim, Djouida. (2024). "Artificial Intelligence from John Searle's Perspective." *Al-Hikma Journal for Philosophical Studies*, 12(3). [In Arabic].
2. Omran, Joudi. (2024). "Artificial Intelligence between Scientific Applications and Philosophical Limits." *Al-Hikma Journal for Philosophical Studies*, 12(3). [In Arabic].
3. Al-Maamari, Saif bin Nasser bin Abdullah. (2024). "Establishing Civil Liability for Artificial Intelligence Applications in Omani Legislation: An Analytical Study." *Al-Ijtihad Journal for Legal and Economic Studies*, 13(2). [In Arabic].
4. Rashid, Nazim Hassan, & Al-Balhad, Mai. (2023). "Auditing Bias in Artificial Intelligence in Light of the Institute of Internal Auditors (IIA) Artificial Intelligence Auditing Framework: A Theoretical Analytical Study." *Journal of Contemporary Business and Economic Studies*, 6(1). [In Arabic].
5. Ben Ali, Zouhir. (2025). "Producing Historical Knowledge in the Age of Artificial Intelligence: A Study of Deep Research Models." *Journal of Humanities and Social Sciences*. [In Arabic].

Foreign References

1. Block, N. (2025). Can Only Meat Machines Be Conscious? *Trends in Cognitive Sciences*, 30(4), 298–308.
2. Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford University Press.
3. Dennett, D. C. (1991). *Consciousness Explained*. Little, Brown and Co.
4. Dreyfus, H. L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. MIT Press.

5. Haugeland, J. (1985). Artificial Intelligence: The Very Idea. MIT Press.
6. Khamassi, M. (2024). Interpreting the Consequence of Chinese Room Argument in AI. HAL Science, 1–12.
7. Nagel, T. (1974). What Is It Like to Be a Bat? The Philosophical Review, 83(4), 435–450.
8. Searle, J. R. (1980). Minds, Brains, and Programs. Behavioral and Brain Sciences, 3(3), 417–424.