Bi-quarterly Journal of
Theosophia Islamica
Islamic Sciences and culture Academy
EISSN: 2783-3240
Vol. 4 ● No. 1 ● 2024 ● Issue 7

Comparative Study of the Capacities of Artificial Intelligence in Reconstructing Human Identity and Consciousness from the Perspective of Mulla Sadra's Transcendent Philosophy and John Searle's Philosophy of Mind

Seyedeh Masoumeh Tabatabaei[1] Arash Rajabi[2]

Received: 2025/04/19 Accepted: 2025/04/27

Abstract

This article provides a philosophical analysis of identity within the context of artificial intelligence, from the perspective of Transcendent Philosophy (especially the views of Mulla Sadra) and contemporary philosophy. With the rapid advancements in artificial intelligence as one of the most significant innovations in computer science, issues such as the reconstruction of human identity and its comparison with natural intelligence have come to the fore. The paper examines the philosophical capacities and limitations of this phenomenon, exploring the perspectives of Mulla Sadra and John Searle, and offers a comparative

1. PhD, Faculty of Theology and Islamic Studies, University of Qom, Qom, Iran (corresponding author).
Email: eshraghm@yahoo.com
2. PhD, Theoretical Foundations of Islam, Islamic Maaref University, Qom, Iran.
Email: eshragh.sadra@gmail.com

* Tabatabaei, S. M. & Rajabi, A. (2025). Comparative Study of the Capacities of Artificial Intelligence in Reconstructing Human Identity and Consciousness from the Perspective of Mulla Sadra's Transcendent Philosophy and John Searle's Philosophy of Mind. *Theosophia Islamica*, *3*(7), pp. 95-114.
https://doi.org/10.22081/jti.2025.71614.1061

analysis of their views on identity and consciousness in machines and artificial intelligence. According to the views of these two thinkers, artificial intelligence, despite its advancements, cannot achieve human identity, as it lacks essential characteristics such as consciousness, intentionality, and abstraction. Searle argues that artificial systems will never be able to attain genuine consciousness because understanding meaning and consciousness are intrinsic to the human brain and mind, qualities that cannot be attributed to machines. He consistently supports this belief through the Chinese Room experiment and various critiques of artificial systems.

Keywords

Artificial intelligence, identity, Transcendent Philosophy, Mulla Sadra, John Searle, philosophy of mind, consciousness.

Statement of the Problem

Artificial Intelligence (AI), as one of humanity's most advanced achievements and a prominent innovation in computer science, began in the mid-20th century with efforts such as the creation of the Turing machine and early projects aimed at simulating human behavior. In recent decades, advancements in machine learning and deep neural networks have transformed AI into a vast and influential field. This technology has not only contributed to solving complex scientific and industrial problems but has also posed new challenges in philosophical discussions on topics such as identity and consciousness.

Among the issues raised in this context are the potential of AI to reconstruct human identity and consciousness. The question arises: could AI one day surpass or match human natural intelligence by acquiring consciousness, or serve as a suitable substitute for it? Is it possible to attribute identity to AI, or not?

This article analyzes and compares the views of Transcendental Wisdom and the perspectives of Mulla Sadra and contemporary philosopher John Searle. The paper aims to offer a comparative examination of the capacities of AI in reconstructing human identity and consciousness from the viewpoints of these two philosophers.

Definition of Artificial Intelligence

Artificial intelligence refers to a branch of computer science that focuses on the design and development of systems and algorithms capable of performing tasks that typically require human natural intelligence. These tasks include natural language understanding, pattern recognition, learning from experiences, decision-making, and

problem-solving. In other words, artificial intelligence refers to the ability of machines to perform tasks that were previously thought to require human thought and reasoning. (Russell, S., & Norvig, P., 2021). In fact, artificial intelligence may potentially serve as a complement to or substitute for human intelligence and decision-making in tasks that require precise pattern analysis, prediction, and assessment of future outcomes.

1. A Philosophical Examination of Human Identity in Mulla Sadra's Thought

1.1. Substantial Motion

In his philosophical system, Mulla Sadra presents "substantial motion" as the foundation for explaining the dynamism and transformation of human identity and the restless nature of the world. He argues that since the natural world is not pure actuality, but rather a display and blend of potentialities and actualities, with motion being the transition of an object from potentiality to gradual actuality, both the nature of the natural world and the essence of human existence are constantly undergoing change and development. This transformation forms the foundation of human identity (Mulla Sadra, 1989, vol. 3, p. 101). From this perspective, artificial intelligence is not considered a natural object or part of the natural world, but rather a product and extension of human existence. Therefore, in its independent form, it lacks substantial motion and cannot be regarded as an entity capable of human-like identity development. Clearly, with the perfection of human substantial motion, achievements such as artificial intelligence—being a product of natural intelligence—will continue to grow and evolve. However, since it lacks an independent essence, nature, and soul, it will not be subject to substantial motion. Even its potential for self-

expansion is based on pre-existing data that human intelligence encodes, rather than the growth and development of its own essence and nature.

1.2. Unity of the Intellect and the Intellected

One of the key principles in Transcendental Philosophy is the unity of the intellect, the intellector, and the intellected (*ittiḥād al-ʿaql wa-l-ʿāqil wa-l-maʿqūl*). This principle asserts that, in the process of cognition, a person becomes one with the object of knowledge. Importantly, humans possess "self-awareness" and have direct, intuitive knowledge of themselves and their essentially known objects. In other words, awareness, the object of awareness, and the identity of the aware individual are united. In Mulla Sadra's philosophy, this is referred to as the unity of intellect, the intellector, and the intellected, or the unity of knowledge, the knower, and the known, or perception, the perceiver, and the perceived (Mulla Sadra, 1989, vol. 3, p. 312). This unique feature of humans creates a fundamental distinction between them and artificial intelligence, which merely processes data. Artificial intelligence does not possess self-awareness and operates solely based on data from natural intelligence. It is not aware of the data processing itself, and thus cannot be considered to have self-knowledge or its data and analyses in the way that humans do, with its knowledge being united with its essentially known object and itself—the knower. In contrast, this philosophical concept is realized in humans according to the framework of Transcendent Philosophy.

Philosophical Analysis

In his philosophy, especially in his theories concerning the primacy of existence, substantial motion, and the unity of the intellect,

the intellector, and the intellected, Mulla Sadra views humans as transcendent and evolving beings who reach their ultimate perfection through substantial motion. According to him, human identity, from both an epistemological and ontological perspective, is not only connected to material dimensions but also to spiritual ones. In other words, humans are "corporeal in origin and spiritual in survival" (Mulla Sadra, 1989, vol. 8, p. 345). Within this framework, humans are beings rooted in the earth and the natural world, but they do not remain there. Through their progressive motion, they continually strive toward perfection, connecting with the divine truth. In this process, their potentialities and capabilities also have the capacity for growth. According to Sadra, humans are not merely material beings; they have the ability to reach the level of abstraction and even transcendence, attaining divine and godly capabilities. This is what the Qur'an refers to as the station of God's vicegerency (*khilāfat Allāh*). Moreover, humans are not simply rational animals, but in the realization of their perfection, they are the most noble type, in which all meanings and perfections are contained and united. Through acquiring knowledge and action, they can ascend to the higher realms and attain the station of *Liqāʾ Allāh* (encounter with God). Sadra believes that the primary material of the human body is clay, which, through development and evolution, takes various forms until it becomes capable of receiving divine grace. When the divine light shines upon the elemental body, another kind of creation occurs. This potential is present in all humans, but in some, with the strengthening of both intellectual and practical faculties, it is actualized through conscious choice and will to attain ultimate perfection (Mulla Sadra, 1989, pp. 129-135). The creation of the universe is for these beings, referred to as the perfect human and the vicegerent of God. Only the perfect human is worthy of the name

"human" and the station of divine vicegerency, embodying the Most Great Name of God and divine perfections. Therefore, from Sadra's perspective, the true identity of humans arises from the earth and, through knowledge and free will, ascends to the heavenly realms, becoming a divine identity that no other being can reach.

Based on this perspective, it is clear that, although artificial intelligence, with its lack of self-awareness and free will, and with its limited capabilities and computational and data structures, may serve as a useful tool in facilitating certain aspects of human life, it possesses a fixed, predefined, and mechanical identity that is imposed by an external agent. According to the philosophy of Mulla Sadra, AI cannot enhance its identity through knowledge and action without the intervention and programming of a human agent. Mulla Sadra believes that humans, in their spiritual development, attain true knowledge. In this view, humans possess specific spiritual dimensions that cannot be confined within data and mathematical algorithms. These qualities, such as connection to the truth and innate understanding, distinguish humans from machines. Mulla Sadra refers to humans as beings capable of receiving the truth through intuition and reason. Therefore, artificial intelligence, which can only process data, is incapable of comprehending the truth in its deeper sense and cannot grasp human identity.

Mulla Sadra is completely opposed to materialistic views. He perceives humans as composite beings, consisting of both body and soul, where the human "soul" or spirit functions as an independent and immaterial essence. Thus, artificial intelligence, which operates fundamentally based on material data and algorithms, cannot attain true consciousness, as this consciousness requires the soul and abstraction.

2. A Philosophical Examination of Human Identity in John Searle's View

John Searle, one of the most prominent contemporary philosophers, has proposed foundational theories in the philosophy of mind and language. He analyzes human identity by focusing on concepts such as intentionality, consciousness, subjectivity of mental states, and biological naturalism.

2.1. Intentionality as the Basis of Human Mental Identity

Searle introduces the concept of intentionality, or the "aboutness" of mental states, as the fundamental characteristic of the mind. Intentionality refers to the mind's ability to refer to objects, states, or matters. This concept, which has also been discussed in phenomenology, according to Searle, is the key distinction between humans and machines. He believes that intentionality is the feature that allows the human mind to refer to something beyond itself, whereas machines lack this ability (Searle, 1983). In this framework, human identity is shaped by the ability to understand, generate meaning, and interact with the external world. However, from his perspective, intentionality is not limited to intentions but also includes beliefs, desires, hopes, fears, love, hatred, greed, disgust, shame, pride, anger, joy, and all conscious and unconscious mental states that refer to or are about the external world (Searle, 2003, p. 24).

2.2. Consciousness: A Necessary Condition for Human Identity

Searle emphasizes the importance of consciousness as a vital element in human identity. He considers consciousness a qualitative and subjective state that is dependent on the internal experiences of

the human being. Searle regards consciousness as the central and fundamental reality of human existence, as without it, human aspects such as language, love, and humor would all be impossible (Searle, 2003, p. 24). Contrary to behaviorist or physicalist theories that reduce consciousness to physical processes, Searle argues that consciousness has an irreducible quality. In his view, consciousness cannot be simply reduced to physical or neural states; it must be analyzed as a mental phenomenon independent of them (Searle, 1992). From this perspective, human identity is not only dependent on brain activity but also on the conscious experiences of the individual.

2.3. Subjectivity of Mental States

The subjectivity of mental states is essentially the distinction between an individual's own perception and that of others. According to Searle, this subjectivity is marked by realities such as the fact that I can feel my pain, but you cannot. I see the world from my perspective, and you see it from yours. I am aware of myself and my internal mental states, which are completely distinct from those of other people and their mental states (Searle, 2003, p. 25).

2.4. Biological Causation: The Foundation of Human Identity

In his theory of "biological naturalism," Searle links human identity to its biological roots. He believes that consciousness and other mental characteristics are products of brain activity, but these activities must be understood in a biological context. In other words, human identity cannot be studied separately from the body or biological structures. Searle emphasizes that this view lies between extreme reductionism and mind-body dualism: mental properties are

real and part of the natural world, but they must be understood in the biological framework (Searle, 2004).

Searle criticizes Descartes' dualistic theories and argues that the mind and body are part of a unified system. He also considers the materialistic views that reduce consciousness to physical states as inadequate. From this perspective, human identity is a combination of mental and biological characteristics that function simultaneously. In John Searle's philosophy, human identity is based on concepts such as intentionality, consciousness, and biology. By rejecting Cartesian dualism and physical reductionism, he presents a comprehensive theory that addresses both mental experiences and biological foundations. Searle emphasizes that these characteristics distinguish human identity from other beings, particularly machines.

John Searle, with his materialistic view of the identity of the soul, believes that mental phenomena are the result of neurophysiological processes in the brain and its characteristics, which he refers to as biological naturalism. He argues that mental events and processes are as much a part of our biological natural history as the stomach, cellular division (both mitosis and meiosis), or enzyme secretion. To explain his view, Searle uses an analogy: Consider water. We are all familiar with the behavior of water at the macro level in everyday objects. We know it is wet, odorless (if pure), drinkable, and takes the shape of its container, among other properties. But why does water behave this way? The answer lies at the microscopic level, where we find that water is made up of millions of invisible molecules, each composed of two hydrogen atoms and one oxygen atom (H_2O). The same principle applies to solid materials. He then provides examples, such as diamond and graphite, both of which

are made of carbon atoms, yet diamond is hard and graphite is soft. In any case, the macroscopic properties are the result of the behavior of the underlying microscopic elements.

Based on John Searle's views in *Minds, Brains, and Programs*, artificial intelligence can simulate human cognitive aspects and achieve similar capabilities to humans. In this approach, machines are capable of making complex decisions and exhibiting behaviors similar to those of humans, based on data and algorithms, although a deep understanding of the meaning of these decisions and their connection to human identity is not possible. In addressing the issue of artificial intelligence, Searle demonstrates that machines lack the key features necessary to form human identity. He argues that human identity depends on the ability to experience intentionality and consciousness, while machines simply process data. Searle uses his famous "Chinese Room" argument to show that machines cannot achieve true understanding, an argument that will be further discussed in this paper.

In his philosophical analysis of artificial intelligence, John Searle distinguishes between two types of AI: Weak AI and Strong AI. This distinction is one of the central debates in contemporary philosophy of mind, and Searle's view plays a pivotal role in critiquing the capabilities of Strong AI.

Weak AI

Searle argues that weak artificial intelligence is merely a simulation of human mental functions. Accordingly, weak AI systems can exhibit behaviors similar to those of humans, but these behaviors do not imply the existence of understanding, consciousness, or a mind within them. In other words, this type of AI is a tool for simulating the

human mind without actually possessing a true mind. Searle states that weak AI can be useful for testing theories about human cognition, but it will never truly understand meanings or possess intentionality (Searle, 1980, p. 417).

Strong AI

In contrast to weak AI, strong AI claims that if a machine can simulate the functions of the human mind, it can be considered to possess a mind, consciousness, and true understanding. Searle strongly disagrees with this view and refutes it based on his "Chinese Room" argument.

Searle's Critique of Strong AI

Searle presents three main criticisms of strong artificial intelligence:

1. Lack of intentionality or "aboutness": Machines lack the ability to refer to external objects or meanings. They only process data (Searle, 1980, p. 418).
2. Lack of conscious experience: Consciousness is one of the fundamental features of the human mind and cannot be easily reproduced in machines.
3. Inability to generate meaning: Machines only manipulate symbols, whereas the human mind is capable of generating meaning.

Searle emphasizes that understanding this distinction is essential to avoid scientific and philosophical misunderstandings. While weak AI can be useful in cognitive sciences and technology development, the claim of strong AI about creating a mind and consciousness in machines is not only exaggerated but also

philosophically incorrect. Searle's perspective on AI, with its distinction between weak and strong AI, provides a philosophical framework for understanding the limitations of this technology. Using the Chinese Room argument, he demonstrates that even the most advanced machines cannot achieve true consciousness and intentionality, underscoring the essential distinction between the human mind and machine functions.

Chinese Room Argument

This famous argument or thought experiment was presented by Searle in 1980, and its purpose was to critique strong artificial intelligence. According to Searle's explanation, imagine a group of computer programmers have written a program that enables a computer to simulate understanding the Chinese language. So, if a question is given to this computer in Chinese, it will match the question with its memory or database and provide appropriate answers in Chinese. To aid the reasoning, assume that the computer's responses are as good as those of a native Chinese speaker. Now, does this computer understand Chinese based on this? In other words, does it understand Chinese in the same way a native Chinese speaker understands it? Well, imagine you are locked in a room, and there are baskets full of Chinese symbols. Suppose you don't know a single word of Chinese, but you are given a book of rules in English for working with these Chinese symbols. These rules specify how to manipulate the symbols entirely formally, in terms of their syntax, not their semantics. So, for example, the rule might say: take a symbol from basket one and place it next to another symbol from basket two. Now, suppose more Chinese symbols are brought into this room, and suppose you are given additional rules for sending symbols out of the room. Imagine that, without knowing it, the symbols brought into the

room are called “questions” by humans outside the room, and the symbols you send out are called “answers to these questions.” Further assume that the programmers have done an excellent job designing the program, and you become very proficient in manipulating the symbols. Soon, your responses are indistinguishable from those of a native Chinese speaker. Here you are, locked in your room, manipulating Chinese symbols and sending out Chinese symbols in response to incoming ones. Based on this situation, there is no way you could learn Chinese merely by working with these formal symbols (Searle, 2003, pp. 60-62).

In short: A person who does not know Chinese is in a room. They have a manual that tells them how to match Chinese symbols in response to specific inputs. It seems that the person in the room knows how to speak Chinese, because they produce meaningful responses. However, in reality, they are simply following symbolic instructions and have no understanding of the Chinese language.

Searle argues that this is similar to how computers operate. They merely execute pre-programmed instructions without truly understanding what is happening. Machines may process information, but there is no awareness or meaning in this process. In fact, strong artificial intelligence cannot simulate the mind. In this thought experiment, Searle assumes that a non-Chinese person is sitting in a locked room and, using a manual, is able to process Chinese instructions and answer Chinese questions. The person in the room has no understanding of Chinese, yet they can still respond correctly to Chinese questions. Searle concludes that such a system (here, a machine) is not capable of "understanding" or "awareness" of the language; rather, it simply follows rules according to predetermined guidelines. In other words, merely matching symbols and signs together cannot equate to awareness and understanding. He argues that

for awareness to exist, a system must genuinely understand meaning, not just operate based on rules (Searle, 1980, p. 417).

Daniel Dennett's Critique of the Chinese Room Argument and Searle's Response

Daniel Dennett has strongly criticized Searle's Chinese Room experiment. In this thought experiment, Searle tried to show that AI systems cannot have consciousness because they only operate by processing symbols and rules, without truly understanding. Dennett disagrees with this idea and believes that understanding and consciousness do not mean awareness of meanings but are simply a functional process that can be attributed based on the intelligent behaviors of a system. In other words, Dennett believes that if a machine can behave correctly like humans, it can be considered conscious, even if it does not truly understand meaning as humans do (Dennett, 1991, p. 195). Dennett further elaborates on this in *Darwin's Dangerous Idea* (1995), where he discusses how mental processes and consciousness can arise from evolutionary processes. He argues that cognitive processes, as observed in humans and other living beings, can be fully modeled and simulated in machines. He believes that if an artificial system can exhibit human-like behaviors, it can be considered conscious, and there is no need for a soul or the concept of deep meaning understanding. In fact, Dennett believes that AI can imitate all the characteristics of the human mind without requiring true consciousness (Dennett, 1995, p. 302).

Daniel Dennett's views on artificial intelligence contrast with those of John Searle, who believes that machine consciousness is impossible. Dennett, emphasizing the theory of new behaviorism and intentional stance, argues that AI systems can behave like humans,

and for this reason, they can experience consciousness and mind in a similar way. According to Dennett, consciousness and intelligence are entirely linked to cognitive and behavioral functions and do not require specific characteristics of humans or living beings. However, Searle believes that human consciousness is inherently tied to the biology of the brain and cannot be reduced to algorithmic processes executed in machines. In other words, machines lack the necessary biological structures to create consciousness, and consciousness is only possible in living systems that possess the appropriate biological structures. Searle argues that strong AI cannot truly create a mind or consciousness because machines are merely symbol processors and lack the ability to understand meaning. He emphasizes the fundamental distinction between the biology of the human brain and the algorithmic processes of machines.

In his Chinese Room theory, Searle emphasizes that even if machines simulate human behaviors, they can never truly understand the real and philosophical meaning of these behaviors. In other words, machines only appear to act like humans but do not understand the substance or meaning behind these behaviors.

Searle has strongly criticized Dennett's views, asserting that although AI can provide simulations of human behaviors, such simulations do not equate to understanding and consciousness. Searle argues that consciousness is an understanding of phenomenological experiences, not merely the processing of information. He believes that consciousness is an irreducible feature specific to biological organisms and cannot be attributed to machine systems. According to him, the human mind has distinct characteristics that set it apart from artificial systems. He refers to this theory as the phenomenology of consciousness and asserts that consciousness is a mental state tied to

subjective experiences, emotions, and the understanding of the meaning of the world. While AI can behave similarly to humans, it cannot possess what Searle calls the real understanding of meaning. He emphasizes that consciousness, as it exists in humans, is a physical phenomenon related to brain function and cannot be fully simulated in machines (Searle, 1992, p. 174).

A Comparative Study of the Views of Mulla Sadra and John Searle

Mulla Sadra views the human identity as evolving through the process of substantial motion. He believes that humans are beings who have emerged from the earth and, through knowledge and free will, reach perfection. This perfection is attained through connection to the divine truth and the manifestation of the divine light in humanity. Therefore, from Mulla Sadra's perspective, human identity is not only material but also spiritual, and it has the potential to attain the state of immateriality (*tajarrud*) and divine vicegerency. Hence, humans possess qualities such as intuitive awareness and abstraction, which artificial intelligence cannot achieve. John Searle also understands human identity in terms of consciousness and intentionality; that is, humans have the capacity to refer to objects and states beyond themselves. Searle emphasizes the distinction between humans and machines in terms of intentionality, arguing that machines are incapable of understanding true meaning or self-awareness. Searle opposes fully mechanical and physical theories of the mind, asserting that the human mind cannot be reduced to mere physical processes or machine-like mechanisms. In contrast to theories such as the computational theory of mind (which likens the human mind to a computer program), Searle states that the mind and consciousness possess qualities that cannot be fully replicated by artificial systems.

Searle is one of the primary critics of AI-based viewpoints that claim consciousness and mind can be attributed to machine-like systems or computer programs. He also believes that the philosophy of mind must go beyond physical and mechanical perspectives, with the human characteristics of the mind being specifically considered.

Searle holds a generally pessimistic view about the future of artificial intelligence. He argues that even in the near future, AI will not be able to attain consciousness. While Searle acknowledges the scientific advancements in AI, he fears that attempts to fully simulate the human mind will ultimately fail because artificial systems will never reach a true understanding of meaning and consciousness. He emphasizes that although machines can be created to behave like humans, they will never be able to possess a mind similar to that of humans (Searle, 1992, p. 196).

Conclusion

In comparing these two viewpoints, it seems that while both philosophers emphasize the distinction between humans and artificial intelligence, Mulla Sadra's focus on the spiritual and transcendent aspects of humanity, including connection to the divine truth and substantial motion, views human identity as an evolved and spiritual being. In contrast, John Searle places more emphasis on the cognitive and conscious dimensions of humanity, highlighting intentionality as the key feature distinguishing humans from machines. Overall, both perspectives imply that artificial intelligence, despite its advancements, cannot attain human identity because it lacks essential qualities such as consciousness, intentionality, and abstraction. John Searle's views on AI and consciousness are particularly notable in opposition to theories by Daniel Dennett and other proponents of

strong AI. Searle argues that artificial systems will never achieve true consciousness because the understanding of real meaning and awareness is a characteristic of the human brain and mind, which cannot be attributed to machines. He has consistently supported this belief through the Chinese Room experiment and various critiques of artificial systems.

References

Dennett, D. (1991). *Consciousness explained*. Little, Brown and Company.

Dennett, D. (1995). *Darwin's dangerous idea*. Simon & Schuster.

Mulla Sadra, M. (2006). *Asrār al-āyāt* (S. M. Mousavi, Trans.). Tehran: Ḥikmat Publications. [In Persian]

Mulla Sadra, M. (2009). *Al-ḥikmat al-muta'āliya fī al-asfār al-arbaʾa al-ʿaqliyya* (Vols. 3 & 8). Beirut: Dār Iḥyāʾ al-Turāth al-ʿArabī. [In Arabic]

Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach*. Pearson.

Searle, J. (2003). *Mind, brain, and science* (A. Divani, Trans.). Qom: Boostan-e Ketab. [In Persian]

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417-424.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.

Searle, J. R. (1992). *The rediscovery of the mind*. MIT Press.

Searle, J. R. (1995). The problem of consciousness. *Journal of Consciousness Studies, 2*(4), 324-332.

Searle, J. R. (2004). *Mind: A brief introduction*. Oxford University Press.