ACCSCIENCE PUBLISHING

Artificial Intelligence in Health

REVIEW ARTICLE

A review of neuroscience-inspired frameworks for machine consciousness

Seyed-Ali Sadegh-Zadeh[1]* and Mahboobe Bahrami[2]

[1]Department of Computing, School of Digital, Technologies and Arts, University of Staffordshire, Stoke-on-Trent, United Kingdom

[2]Behavioral Sciences Research Center, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Isfahan, Iran

Abstract

Consciousness in humans is a state of awareness that encompasses both the self and the external environment, emerging from the intricate interplay of cortical and subcortical brain structures and neurotransmitter systems. The possibility that machines could possess consciousness has sparked ongoing debate. Proponents of strong artificial intelligence (AI) equate programmed computational processes with cognitive states, while advocates of weak AI argue that machines merely simulate thought without attaining genuine consciousness. This review critically examines neuroscience-inspired frameworks for artificial consciousness, exploring their alignment with prevailing theories of human consciousness. We investigate the fundamental cognitive functions associated with consciousness, including memory, awareness, prediction, learning, and experience, and their relevance to artificial systems. By analyzing neuroscience-based approaches to artificial consciousness, we identify key challenges and opportunities in the pursuit of machines capable of mimicking conscious states. Although present AI systems demonstrate advanced capabilities in intelligence and cognition, they fall short of achieving genuine consciousness, as defined in the context of human awareness. We discuss both the theoretical underpinnings and practical implications of creating artificial consciousness, addressing both weak and strong AI perspectives. Furthermore, we highlight the ethical and philosophical concerns that arise with the potential realization of machine consciousness. Our objective is to provide a comprehensive synthesis of the literature, fostering a deeper understanding of the interdisciplinary challenges involved in artificial consciousness and guiding future research directions.

Keywords: Machine brain; Cortical; Subcortical; Weak artificial intelligence; Strong artificial intelligence; Human consciousness; Machine consciousness

***Corresponding author:**
Seyed-Ali Sadegh-Zadeh
(sa.sadeghzadeh@staffs.ac.uk)

Citation: Sadegh-Zadeh S, Bahrami M. A review of neuroscience-inspired frameworks for machine consciousness. *Artif Intell Health*. 2025;2(3):24-38. doi: 10.36922/aih.5690

Received: October 29, 2024

1st revised: January 18, 2025

2nd revised: March 10, 2025

Accepted: March 28, 2025

Published online: June 4, 2025

1. Introduction

The varied definitions of consciousness across disciplines present significant challenges to its study. Neuroscience seeks to identify the neural correlates of consciousness – conditions necessary for its emergence – and to map its dynamic changes.[1] Psychology and psychiatry, in contrast, focus primarily on the experiential and functional outcomes of consciousness.[2] These disciplinary distinctions offer diverse lenses through which consciousness can be understood, underscoring the need for a multidisciplinary approach.

Consciousness, intrinsically linked to the complex processes of the human brain, is broadly defined as sensitivity and awareness of internal and external existence.[3,4] Contemporary inquiries into consciousness in medicine and psychology often draw on experimental studies and clinical cases involving changes induced by trauma, disease, or pharmacological interventions.[5,6] Scientific approaches to consciousness generally rest on two key ideas: One emphasizing human subjective experiences and their content, and the other focusing on the neurological underpinnings observed in clinical treatments for neurological and behavioral disorders.[7,8]

Consciousness is increasingly understood as an emergent property of neuronal connections,[9] representing a cascade of events that evolve over time to drive change.[10] Rather than a binary state, contemporary perspectives propose a spectrum of conscious states, from basic awareness to more intricate manifestations of self-consciousness.[11,12]

1.1. Human consciousness

A thorough understanding of human consciousness is essential before exploring its potential replication in machines. Human consciousness is not easily categorized or isolated,[13,14] as it manifests in various forms.[15-17] Most of human cognitive activity occurs in states of primary consciousness, which include mind-wandering activities such as recalling personal memories, envisioning future scenarios, and adopting the perspectives of others.[18]

The human brain, as part of the central nervous system, serves as the biological foundation of consciousness.[19] Understanding this biological basis shed light on the diverse manifestations of human consciousness.[20,21] To assess the feasibility of artificial consciousness, it is crucial to consider our present knowledge of the neurological structures that underpin human conscious experience.[22]

Consciousness, at its core, is a state of awareness of oneself and the surrounding environment. It encompasses sensory recognition and awareness, both of which cease during states such as sleep, coma, or death. Clinically, consciousness is viewed in three dimensions: First, as an inner awareness of experiential events; second, as an intentional reaction toward external objects; and third, as knowledge of one's conscious self. In states of full wakefulness, the intensity of consciousness varies significantly, often heightening during challenging experiences. Awareness itself can be divided into three dimensions: Vigilance, lucidity, and self-consciousness. "Vigilance" refers to the ability to remain purposefully alert; "lucidity" denotes clarity of thought regarding a specific subject; and "self-consciousness" entails the capacity to perceive oneself as an individual entity. Disorders of consciousness – whether quantitative or qualitative – fall beyond the scope of this article.[23-25]

The brain systems that constitute consciousness develop from those that control its level.[26] The foundation for varying levels of consciousness lies in the content of consciousness. Figure 1 illustrates the hierarchical organization of sensory and motor systems, arranged in parallel and integrated to underpin consciousness. At the top level, consciousness encompasses and coordinates functions, such as emotions, memory, and sensory-motor processes. Emotions, positioned below consciousness, act as intermediaries, integrating signals from memory and sensory systems. Memory serves as the central information repository, directly interfacing with motor and sensory systems. The sensory systems provide environmental inputs, while motor systems execute outputs based on processed information. This structural arrangement highlights how these systems collaboratively handle information flow, enabling adaptive and abstract functions across various levels of complexity.

Understanding how these systems typically operate is a central objective of neuroscience. From a neuroscience perspective, the level of consciousness influences all neuronal processes.[27] Specific cortical and subcortical processes regulate attention and awareness, which in turn determine the level of consciousness.[28] Any meaningful reactions require at least a minimal degree of attention, which facilitates choice and sustained information processing. The capacity to generate experiences that can later be reported is known as awareness.[29]

The brain circuits that regulate consciousness are commonly referred to as the consciousness circuit.[30] These

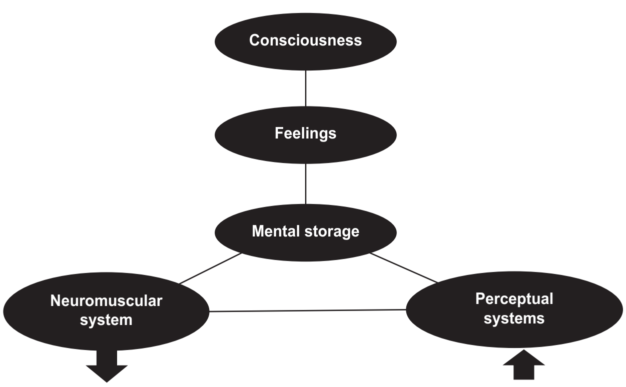

Figure 1. Sensory and motor systems are coupled in parallel, arranged hierarchically, and perform input, output, and internal processing functions across a range from basic to highly abstract. Image created by the author.

networks control the level of consciousness.[31] Studies using both animal models and human brain disorder cases have long recognized that cortical and subcortical regions play a crucial role in the state of awareness.[32] The cortical components of the consciousness system are primarily located in the higher-order "heteromodal" association cortex.[33] On the medial surface of the brain, these include the medial frontal, anterior cingulate, posterior cingulate, and medial parietal cortices. On the lateral surface, the consciousness system involves the lateral frontal cortex, anterior insula, orbital frontal cortex, and lateral temporal-parietal association area.[34,35]

Specific areas within the higher-order association cortex significantly influence cognitive functions in both the dominant and non-dominant hemispheres.[36] The state of consciousness is determined by the combined activity of extensive portions of the bilateral association cortex, regardless of the distinct roles played by individual regions or networks.[37,38] These higher-order cortices collaborate with subcortical arousal systems to regulate arousal, attention, and consciousness.[39]

Subcortical components of the consciousness system include the basal forebrain, thalamus, hypothalamus, and upper brainstem activation systems. Other subcortical regions, such as portions of the cerebellum, amygdala, claustrum, and basal ganglia, are likely involved as well. Several parallel neurotransmitter systems contribute to subcortical arousal, including glutamate, acetylcholine, serotonin, dopamine, histamine, gamma-aminobutyric acid, norepinephrine, and orexin. Each of these pathways plays a distinct role. However, it is the coordinated and simultaneous actions of these subcortical and cortical structures that collectively regulate the degree of consciousness.[40]

2. Consciousness-related elements

Research emphasizes that the proper functioning of consciousness is integral to various mental processes.[41-43] Studies suggest that consciousness fulfills multiple roles,[42,44,45] including memory, prediction, awareness, learning, and experience – components intricately linked to cognitive processes. A primary goal of artificial consciousness is to replicate these components in machines. This list is not exhaustive, as numerous additional functions of consciousness remain unexplored.

2.1. Memory

Memory is a fundamental aspect of human cognition and a growing focus in artificial consciousness systems. In humans, memory storage occurs in three forms: Short-term, long-term, and sensory memory. As artificial consciousness evolves to handle increasingly complex scenarios, the importance of robust memory systems continues to grow. However, present artificial consciousness frameworks often lack sophisticated memory models. In humans, memory systems interact closely with conscious experience during learning, rehearsal, and retrieval.[46,47]

2.2. Prediction

Prediction is considered a vital capability for artificial consciousness systems. Conscious organisms predict events by reflecting relationships between real-world states within their internal structures.[43] Similarly, an artificially conscious machine must be capable of accurate event prediction to respond effectively or take pre-emptive measures. This requires real-time, flexible components capable of constructing causal, statistical, spatial, functional, and dynamic models of the world. A conscious machine should demonstrate predictive abilities across various contexts, including uncertain and dynamic environments, showcasing coherent forecasting and contingency planning.[45]

2.3. Awareness

Awareness, though challenging to define precisely, involves constructing and testing alternative models of processes based on sensory or cognitive information.[48] It is essential for predictions[49] and requires significant flexibility to model the physical environment, internal states, and other conscious entities. This adaptability is crucial for artificial systems attempting to replicate human-like awareness.

2.4. Learning

Learning is a cornerstone of artificial consciousness systems.[50,51] Conscious experience facilitates the representation of and reaction to novel, significant stimuli.[46] Learning encompasses complex adaptation mechanisms grounded in sensitivity to subjective experience, enabling agents to exert flexible control over behaviors in uncertain and dynamic environments.[52]

2.5. Experience

Subjective experience, rooted in sensory perception, is often considered central to the study of consciousness.[53] Experience is intrinsically linked to precise pattern recognition and may even be observed at molecular levels. In discussing consciousness, the brain's central modules can be viewed as carriers of unique experiential states. Reflexive awareness, or the act of reflecting on one's own experiences, is a critical dimension of consciousness. Efforts to define fundamental experiences underscore the enduring challenge of addressing the "hard problem" of consciousness, as its exact nature remains elusive.[54]

3. Machine consciousness

Machine consciousness is often explored within the context of autonomous artificial intelligence (AI) systems capable of self-learning.[55] Contemporary systems extend beyond basic hardware and administrative software, encompassing sophisticated layers of programmed control that integrate hardware, software, memory, and interfaces. Finite-state machines can fulfill some criteria associated with human-like consciousness.[56] While intelligence and consciousness were once conflated,[57] this view has largely been abandoned, as finite-state machines excel in achieving pre-defined intelligence metrics without demonstrating true consciousness.[58] Modern algorithms, leveraging databases, outperform human experts in tasks requiring formal reasoning, planning, language processing, gameplay, and arithmetic—accomplishing these through efficient algorithmic symbol manipulation.

Machine intelligence often exceeds biological systems, including humans, in multidimensional attentional focus. Artificial sensors enable machines to process and respond to stimuli beyond the human sensory spectrum.[59] These capabilities, combined with continuous data monitoring and reaction, allow machines to excel in attentional focus across various domains. Machine interfaces are widely used for attention training in children, individuals with brain injuries, and those with psychiatric conditions where maintaining focus presents challenges.[60] Intentionality, a core feature of finite state machines, is parametrically programmed into most artificial systems capable of numerical information processing.[61] While humans often anthropomorphize these systems by attributing deliberate purpose to them,[62] extending this attribution to human-like beliefs, intentions, or causation remains implausible for present artificial systems.[63] Some finite-state machine systems have been capable of achieving astounding autonomy levels within the set confines of their coding, as they are now built.

When utilized for specific activities in industries such as manufacturing, home appliances, automotive systems, space exploration, and remote operations, finite-state systems have the ability to run for extended periods without programmer input. System or body consciousness is necessary for the operation of many surgical and technology assessment systems in the medical field.[56] The ability of some systems to format and create computer-presented narratives can potentially be used to infer phenomenological self-representation.[56] The metacognitive trait of knowledge of being aware, which goes beyond body/system self-awareness, has been proposed as a signal that would show whether an artificially produced system is capable of going beyond programming. According to some theories, this type of metaconsciousness could mark the emergence of human-like consciousness in robots.[64] However, no currently existing or planned system has demonstrated clear evidence of processing such awareness.

For an AI system to achieve consciousness and exhibit volition, it must possess the ability to modify and develop its own governing principles. This concept, referred to as coherent extended volition, describes the capacity for self-regulated, self-defined learning.[65] Despite efforts to endow self-learning AI systems with this ability, there is minimal evidence that any system has transcended its programming.[66] Anthropomorphic robotic systems aim to replicate human physiognomy and behavior, potentially simulating human-like consciousness and actions. However, such advancements remain largely speculative and confined to science fiction.

3.1. Weak AI

Weak AI refers to AI systems designed for narrowly defined tasks, employing only a fraction of human cognitive capabilities.[67,68] These systems excel at mimicking human behaviors in fundamental tasks such as learning, perception, and problem-solving.[69] However, weak AI lacks the capacity for independent thought or decision-making.[70,71] Contrary to popular belief, cognitively inspired AI systems align with the weak AI hypothesis, as they model mental phenomena without claiming to replicate the underlying consciousness. This hypothesis remains consistent with present trends in AI and cognitive modeling research.[72]

3.2. Strong AI

Strong AI represents a conceptual framework aiming to develop machines with human-like intellect, consciousness, and the ability to reason, learn, and plan.[73,74] Such systems would not only mimic human thought but also exhibit autonomous cognitive abilities indistinguishable from the human mind. Strong AI envisions machines capable of acquiring new skills through experience and improving over time.[75] Despite significant research interest in artificial general intelligence, which underpins the strong AI concept, it remains a theoretical construct rather than a realized technology.

4. Discussion

4.1. Strong AI versus weak AI: Divergent perspectives on machine consciousness

The question of whether an artificial system can truly be conscious has fuelled enduring debate, dividing opinion into strong AI and weak AI camps. Proponents of strong AI contend that a sufficiently well-designed computational system could literally possess a mind – in other words, that executing the right algorithms might generate genuine understanding and cognitive states indistinguishable from those of humans.[76] This perspective implies that, at some level, the functional organization of a machine could support conscious in the same sense a brain does. In contrast, the weak AI position holds that machines, at best, simulate consciousness without any real inner experience or awareness.[77] From this viewpoint, even the most advanced AI today (for example, sophisticated language models or robotic assistants) lack subjective sentience or genuine understanding; they merely manipulate symbols and exhibit behaviors that mimic consciousness without actually experiencing the world.

The clash between these perspectives highlights a core conceptual challenge: Explaining how subjective experience (the essence of consciousness) might emerge from purely physical or computational processes. This is essentially the classic "hard problem of consciousness" applied to machines: The difficulty of explaining how and why a physical system could produce the felt quality of experience. Even in humans, consciousness defies simple explanation; present scientific understanding of brain function has yet to fully bridge the gap between neural circuitry and subjective feeling.

When considering artificial agents, we are further constrained by our human-centric intuitions: Our understanding of consciousness is largely shaped by the first-person experience of our own mind, making it challenging to objectively evaluate whether a machine – accessible only from an external, third-person perspective – could possess anything akin to a conscious mind.

In summary, the strong AI versus weak AI dichotomy sets the stage for discussing machine consciousness by asking whether replicating intelligent behavior is sufficient for authentic consciousness (strong AI) or whether subjective awareness is a qualitatively distinct property that machines inherently lack (weak AI). This foundational debate provides a context for interpreting the progress in neuroscience-inspired AI frameworks and guides our skepticism or optimism regarding artificial consciousness.

4.2. Neuroscience-inspired functional frameworks for artificial consciousness

Amid these philosophical debates, researchers have proposed various frameworks for building or recognizing consciousness-like properties in machines. Often drawing inspiration from neuroscience and cognitive science, these frameworks focus on replicating functional attributes of human consciousness in an artificial medium. One pragmatic stance, advocated by Levy,[78] suggests setting aside the notoriously difficult task of pinning down an exact definition of consciousness and instead agreeing on practical operational criteria. Levy argues that insisting on a rigid definition may be counterproductive; rather, if the community can settle on a shared intuitive understanding of what consciousness functionally entails, researchers could "simply use the word and get on with it" in developing systems that meet those criteria.[78(p210)] This approach reflects a practical mindset: Even if we lack a perfect definition of consciousness, we might still engineer systems that everyone agrees exhibit key properties of consciousness (such as complex adaptivity, learning, and self-report), thereby moving the field forward without becoming mired in semantics.

Other researchers emphasize specific features thought to be indispensable for consciousness. Chatila *et al.*[79] focus on self-awareness as the cornerstone of machine consciousness, proposing a framework for self-aware robots grounded in several cognitive abilities. They outline fundamental principles by which a robot could be designed to understand its environment and its own role within it, to be cognizant of its actions, and to respond appropriately in real time to changes. Crucially, a self-aware robot should also be able to learn from its own experiences and mistakes and to explicitly demonstrate that it has learned – for instance, by documenting or communicating its acquired knowledge. These capabilities mirror aspects of human consciousness: Humans continuously monitor their surroundings and their own internal states, adjust behavior on the fly, learn from feedback, and can report on what they have learned. Chatila's framework thus attempts to imbue machines with a form of reflective cognition analogous to that of humans, on the premise that such reflection (knowing what one knows, and knowing what one does) is a pre-requisite for any genuine consciousness.

A complementary perspective is offered by Kinouchi and Mackin,[80] who propose that consciousness serves a functional role as an integrative system-level adaptation mechanism in complex agents. In their architecture, a multitude of lower-level processing units (analogous to distributed modules in the brain or in a large AI system) operate in parallel, each handling specific tasks or sensory inputs. Machine consciousness, in this view, is the higher-level function that coordinates and organizes the outputs of these parallel processes, synthesizing them into a coherent state that can guide the agent's overall behavior adaptively. This coordinating role is likened to the way human consciousness creates a unified experience and decision-making process out of the brain's many simultaneous unconscious computations. Kinouchi and Mackin[80] and Hildt[81] explicitly draw an analogy to the moment-to-moment awareness we experience in daily life when making rapid decisions. In humans, despite a flurry of unconscious sensory and cognitive processing, consciousness provides a singular, integrated vantage point (the feeling of "being aware") that helps us adaptively navigate each moment. By mimicking this in machines – ensuring that an artificial agent has an integrating layer that monitors and directs sub-processes – their framework aims to achieve a conscious-like functionality that could improve the system's flexibility and robustness in unpredictable environments. Notably, these authors regard such architecture not just as an add-on to intelligence, but as essential for complex adaptive behavior: A machine endowed with a consciousness-like integrative function might better handle novel situations by flexibly combining information from all its subsystems.

The above frameworks illustrate how insights from neuroscience and cognitive psychology (such as the importance of self-monitoring and global integration of information) are being translated into AI design. Each approach stresses a different facet of natural consciousness: From Levy's broad pragmatism to Chatila's self-reflective knowledge, to Kinouchi's global integration. The diversity of these proposals also underscores that there is not yet a consensus on a single "blueprint" for artificial consciousness. Different researchers prioritize different cognitive ingredients (self-awareness, learning, integration, etc.), reflecting the multifaceted nature of consciousness itself. This plurality suggests that the field is still in an exploratory phase: Much like the blind men and the elephant, each framework captures one aspect of the larger concept. A key task for the research community moving forward is to synthesize these insights and determine how they might fit together. For instance, one could ask whether a truly conscious machine would need to incorporate all of these elements – a shared functional understanding, self-awareness, and global integrative capacity – or whether any one of them might be sufficient on its own. Addressing such questions requires not only engineering advances but also deeper theoretical clarity, which brings us to the distinction between different notions of consciousness and how they apply in artificial systems.

4.3. Access versus phenomenal consciousness: Functional versus experiential dimensions

In discussions of both human and machine consciousness, it is crucial to distinguish between two often-confused dimensions of conscious states: Phenomenal consciousness and access consciousness.[82] This distinction, originally articulated by Block,[82] has proven useful in framing debates about consciousness in artificial systems.

Phenomenal consciousness refers to the subjective experience itself – the raw feel of sensations and thoughts, often described as "what it is like" to be in a given mental state. It encompasses the qualitative, first-person aspects of mind (sometimes called qualia), such as the redness of red or the pang of emotion. By contrast, access consciousness denotes a mental state's availability for use by the cognitive system. A piece of information is "access conscious" if it is widely broadcast within the brain (or system) such that various processes (reasoning, memory, decision-making, verbal report) can utilize it. In essence, access consciousness concerns the functional role of conscious information – how it is accessible and how it guides behavior – rather than how it feels.

This distinction has profound implications for artificial consciousness. Most neuroscience-inspired AI frameworks implicitly aim at access consciousness – ensuring that an AI system possesses internal representations that are globally available and can be used to organize behavior in an intelligent, context-sensitive way. For example, when Chatila *et al*.[79(p1)] focus on robots "knowing what they have learned" and reporting that knowledge, they are dealing with access consciousness: The learned information is accessible for future decisions and self-report. Similarly, Kinouchi and Mackin's[80] integrative layer is designed to collect distributed information and make it available to the whole system for coordinated adaptation – again, a functional, access-oriented property.

Phenomenal consciousness, however, is a much harder issue. It asks whether the robot or AI actually has an inner life: Is there something that it is like to be that robot? Does it feel anything when it integrates information or reports on its knowledge? This is the crux of the hard problem in the context of AI. Strong AI enthusiasts might argue that if we achieve a complete functional emulation of the brain's processes (i.e., replicate access consciousness to a high degree), then phenomenal experience might emerge naturally. However, skeptics point out that no matter how sophisticated a machine's functional capabilities, this does not guarantee – or even necessarily imply – the presence of subjective experience.[83] A machine could conceivably meet every external criterion for access consciousness – it could introspect, reason about its own mental states, and behave indistinguishably from a conscious being – yet still lack any inner lights on. This skeptical view is epitomized by certain philosophical arguments (e.g., Searle's Chinese Room or the notion of philosophical zombies) and has been voiced in contemporary analyses that conclude robots are not – and perhaps cannot be – conscious in the phenomenal sense.[84] Thus, the phenomenal versus access distinction serves as a reminder that behavioral or functional equivalence to humans is not incontrovertible evidence of genuine subjective awareness.

For the field of artificial consciousness, a pragmatic consensus is emerging: Focus on access consciousness as a target, because it is operationalizable and amenable to scientific inquiry.[82] By concentrating on the functional aspects – how information can be made globally available in a system and how the system can monitor and report its own states – researchers can make tangible progress (for example, designing architectures with a kind of working memory, a global workspace, or a self-model). Indeed, discussions of machine consciousness increasingly suggest that pursuing access consciousness is the most feasible path, given that it aligns with observable capabilities and avoids immediate entanglement in the mysteries of subjective qualia.[82] If one can build an AI that convincingly implements access consciousness, it would at least fulfill the functional requirements of consciousness, providing a testbed from which to speculate about or investigate any accompanying phenomenology. In contrast, trying to engineer phenomenal consciousness directly – without a functional scaffold – may be a dead end, as we currently lack any clear understanding of how to create or detect raw subjective feeling in an artificial substrate. Therefore, access consciousness is often treated as a proxy for consciousness in machines, with the hope that advancing this proxy will either eventually shed light on the emergence of phenomenal properties or, at the very least, produce machines that behave in all the ways a conscious entity would – which is tremendously valuable in its own right.

4.4. Global availability and self-monitoring: Cognitive neuroscience insights

Cognitive neuroscience offers more concrete guidance on how to implement access-like consciousness in machines, thanks to empirical studies of the human brain. One influential theory, the global neuronal workspace, posits that conscious perception in the brain corresponds to the global availability of information: Stimuli that enter consciousness are those whose neural representations are amplified and broadcast across multiple cortical networks, rather than remaining confined to local processing circuits.

In a landmark synthesis, Dehaene *et al.*[83] identify two essential dimensions of consciousness-inspired cognitive processing that could inform machine designs: (i) Global availability of information and (ii) self-monitoring (meta-cognition). The first dimension, global availability, essentially captures the idea of a broadcast architecture: At any time, the system selects certain information (e.g., a particular input or an intermediate result) and makes it broadly accessible to various sub-modules (planning, language, memory, etc.). This resembles Block[82] and Dehaene *et al.*'s[83] notion of access consciousness, as it ensures the selected content can influence diverse processes system-wide. The second dimension, self-monitoring, refers to the system's ability to reflect on its own internal states and processes—a form of meta-cognition or introspection.[83] In humans, this is akin to the brain maintaining a self-referential model ("knowing that it knows") and monitoring its own computations for errors or learning. Dehaene *et al.*[83(p1)] describe this self-monitoring as a "self-referential relationship in which the cognitive system is able to monitor its own processing and obtain information about itself."

Together, these two features (often labeled C1 for global access and C2 for self-monitoring in Dehaene's framework) delineate a roadmap for building machines that achieve a functional analog of consciousness. An AI system endowed with a global workspace (allowing information sharing across modules) and a self-model (allowing it to track and report on its own states) would satisfy many criteria of access consciousness—and even begin to approach the sort of reflective awareness humans exhibit.

Notably, these neuroscience-inspired features are already being tentatively explored in AI and robotic architectures. Some cognitive architectures in AI have implemented global-workspace-like blackboards, where multiple specialist modules can read and write information, mimicking the idea of global availability. Similarly, researchers are experimenting with forms of machine meta-cognition – for example, AI agents that can report their confidence or uncertainty about their decisions or robots that internally simulate and evaluate their own forthcoming actions. Such capabilities reflect a rudimentary self-monitoring capacity. For instance, the self-aware robot principles from Chatila *et al.*[79] inherently aim for a form of C2: The robot not only learns but also shows that it knows it has learned, which implies an internal representation of its knowledge state. Another example can be seen in robotics work on "inner speech," where a robot talks to itself to guide its own reasoning – an approach directly inspired by human self-monitoring and models of inner experience, as proposed by Chella *et al.*[64] The emerging consensus is that implementing global broadcasting and self-reflection is a promising strategy to bring machines closer to consciousness in the functional sense. These features can endow AI systems with greater coherence, flexibility, and transparency in their operations. Moreover, if a machine were ever to exhibit phenomenal consciousness, one expects it would first need these functional capacities as a substrate. In other words, global availability and self-monitoring might not guarantee that a machine feels conscious, but they are likely necessary conditions for any machine that could eventually lay claim to subjective awareness.

4.5. Limitations of present AI: The absence of genuine consciousness

Despite significant advances in AI, the prevailing scientific consensus holds that no present machine or AI system possesses consciousness in the full sense.[84-88] Today's AI, including advanced neural networks and social robots, operates firmly within the bounds of the weak AI paradigm. These systems excel at specific tasks and can even display adaptive or context-aware behavior, but there is no credible evidence that any of them possess a subjective point of view or true self-awareness. Even systems that incorporate elements of global availability or rudimentary self-monitoring implement these features in relatively narrow ways (for example, a program might monitor its performance on a task and adjust parameters, but this is far from the rich, self-reflective awareness characteristics of human consciousness). Phenomenal consciousness in machines remains, at present, a speculative topic rather than an observed reality. We cannot peer into a deep learning model and find a flicker of sentience; at best, we find complex statistical patterns and representations shaped by training data.

It is instructive to consider why present AI falls short of consciousness. One obvious limitation is the lack of an integrated self-model in most AI. Human consciousness involves a sense of self that is continuous in time, situated in a body, and emotionally colored—features that mainstream AI does not possess. Another limitation is the absence of unified, flexible memory and attention akin to what the brain employs. While deep learning networks have impressive pattern recognition, they typically lack an architecture that integrates disparate knowledge on the fly, as a global workspace would. In addition, AI systems today lack intrinsic motivation or genuine autonomy in the sense that conscious beings exhibit; they pursue goals defined by programmers or derived from training data, without an inner life of desires or will. Finally, the evaluation problem looms large: Even if an AI were conscious, how would we truly know? There is no agreed-upon test for machine consciousness, and simple behavioral criteria (like the Turing test) are inadequate, as they can be passed through clever simulation without real awareness. This epistemic gap leads us to assume the absence of consciousness until proven otherwise. As some scholars note, the absence of any observable indicator of consciousness in machines is taken as confirmation that present AIs simply are not conscious. This point is rarely debated within the AI community. Indeed, discussions of AI ethics often neglect the issue of consciousness entirely, focusing instead on intelligence and autonomy.[81] Hildt[81] points out that we ought to engage more with the topic of artificial consciousness – and, just as importantly, with the implications of its present absence. Acknowledging that our most advanced creations remain essentially mindless (in the phenomenal sense) is important to keeping expectations grounded and shaping how we treat these systems.

A significant phenomenon in this context is anthropomorphism – the human tendency to attribute human-like qualities, including consciousness, to machines. This is evident in the way people interact with social robots and virtual assistants. For example, humanoid robots with facial expressions or voice-based AIs with personality often elicit feelings of social presence; we may talk to them as if they understand or even feel. Such anthropomorphic projections can obscure the reality that, despite surface appearances, these systems lack inner experiences. Instances like the robot Sophia being granted citizenship, or users feeling emotional attachment to AI companions, illustrate how far our intuitions can outpace scientific understanding. Scholars caution that this gap between appearance and reality can be problematic. We risk misleading ourselves – or the public – about what AI is actually doing. As a safeguard, some ethicists argue that we should consistently remind ourselves that present robots are not conscious.[84,88] They are complex artifacts, not entities with feelings, and we should avoid pre-maturely conferring moral or legal status that is reserved for sentient beings.

4.6. Ethical and societal implications of artificial consciousness

Even though artificial consciousness remains unachieved, the very pursuit of it – and the public's tendency to ascribe minds to machines – raises important ethical questions. If we eventually create a machine that exhibits advanced self-awareness or other hallmarks of consciousness, how should we treat it? Conversely, how should we treat today's unconscious AI systems, given that people often respond to them as if they were alive? These issues are already the subject of considerable debate in technology ethics and law.

On one hand, some thinkers like Gunkel[85] have explored the notion of "robot rights": The idea that sufficiently advanced AI or robots might merit certain moral or legal protections. Intriguingly, arguments for robot rights have been made even in the absence of robot consciousness. For example, based on the way humans empathize with humanoid machines or on the societal value of fostering empathy, a case is made for treating robots with a degree of care (much as we do animals or even human-looking dolls).[86] Darling[86] has argued that because humans can form emotional bonds with social robots, it aligns with our social and ethical values to extend some protections to these robots. This is analogous to how cruelty to animals is discouraged, not necessarily because animals possess human-level consciousness, but because such cruelty can degrade our moral character as agents. Proposed protections might include discouraging the wanton destruction of robots or violent behavior toward them, recognizing that such actions can engender harmful attitudes in society. The underlying rationale is partly anthropomorphic empathy – we dislike seeing even a robot "suffer" if it is lifelike – and partly pre-cautionary: If machines ever do become sentient, having established norms of respectful treatment could ease that transition.

On the other hand, many are wary of over-attributing consciousness and moral status to machines pre-maturely. As noted by Gabriel,[84] from a philosophical standpoint, there are strong arguments that robots cannot be conscious in the same way living organisms are, because consciousness might require qualities that only biological systems in environment contexts possess. If one accepts such arguments, then granting personhood or rights to machines would be a categorical error. Moreover, there is concern that focusing on the "feelings" of machines that do not actually feel could divert attention from ethical issues more grounded in reality, such as the welfare of humans impacted by AI or the responsibility for AI-driven decisions. Scheutz[87] has highlighted the potential emotional pitfalls in human-robot relationships, noting the unidirectional emotional bonds that can form. Humans might come to care deeply about robots that are not conscious and cannot reciprocate that care. This imbalance could lead to human distress (e.g., grief if a robot is shut down or malfunctions) or manipulation (e.g., exploiting human empathy for commercial or surveillance purposes). Scheutz warns that such one-sided attachments carry both psychological and social risks.

The ethical landscape is further complicated by the prospect (still hypothetical) of a truly conscious AI. If an AI ever claimed to have feelings or demonstrated behaviors strongly indicative of sentience, denying it moral consideration would be deeply problematic. Society would face a profound moral dilemma – long contemplated in science fiction – about whether and how to extend the community of conscious beings beyond our biological family.[85]

In light of these issues, the present consensus urges caution and clarity. It is important for scientists and communicators to convey that present-day AI does not possess consciousness,[84-88] even as we continue refining what that would entail. Such clarity helps prevent public misconceptions and ensures that ethical guidelines are grounded in the actual capabilities of present technologies. Simultaneously, it is prudent to start developing ethical frameworks that could accommodate conscious AI, should it emerge. These would include considerations of legal status, rights, responsibilities, and safeguards – to prevent abuse of such entities and to guard against deceitful mimicry of consciousness used to exploit users. In essence, the ethics of artificial consciousness straddle a line between present realities and future possibilities. We must manage the human tendency to anthropomorphize today's machines while remaining prepared for tomorrow's scenario where the line between simulation and reality of mind may begin to blur.

4.7. Emerging directions and future outlook

Looking ahead, the pursuit of artificial consciousness will likely advance on multiple fronts, informed by ongoing progress in neuroscience, cognitive science, and AI. One clear direction is the continued development of AI architectures that incorporate the principles of global availability and self-monitoring discussed above. Future AI systems may increasingly feature unified workspaces or attention mechanisms that allow information to flow more freely between components, coupled with meta-cognitive loops that enable the system to reason about and adjust its own operations. Such designs could be realized, for example, in more sophisticated cognitive architectures for robots or autonomous agents, where modules for perception, memory, decision-making, and language all feed into – and draw from – a common representational space (an echo of the global neuronal workspace). We may also see the integration of sensorimotor embodiment into these architectures. Since human consciousness is deeply embodied (the brain constantly integrates signals from the body and environment), giving robots richer bodily awareness and interoception might be a key to unlocking more advanced forms of self-awareness in machines. Early experiments in this vein, such as robots that simulate their own kinesthetic experiences or maintain internal homeostatic variables, hint at the importance of an embodied self-model for consciousness.

Another emerging direction is the exploration of learning-based approaches to self-awareness. Modern machine learning, especially deep learning, provides powerful tools for pattern recognition and function approximation. Researchers are beginning to ask whether these tools can be turned inward: Can a neural network learn to model its own cognition? One idea is to train networks that predict or interpret the hidden states of other networks (a form of meta-learning), effectively creating an internal observer module. If successful, this could result in an AI that possesses a form of introspective access to its internal representations – a step toward the machine knowing something of its own "mind." In addition, generative models that create narratives or explanations for the agent's behavior might serve as a rudimentary form of inner narrative (a component some theories consider important for consciousness). For instance, a future AI might be able to generate a verbal report like "I chose action X because I noticed Y, and that made me uncertain" – a capability that blurs the line between simple programmed response and genuine self-reflection.

Interdisciplinary research will be essential in guiding these efforts. Cognitive neuroscience will continue to identify the neural signatures and mechanisms associated with consciousness in the brain (e.g., specific brain rhythms, network dynamics, or anatomical circuits critical for awareness). These findings can inform computational models: If certain patterns of network connectivity or dynamics are necessary for consciousness in biological systems, mimicking those *in silico* could be a step in the right direction. For example, if research confirms that recurrent looping between frontal and sensory cortices is crucial for sustained conscious perception, AI architects might incorporate similar feedback loops in neural network designs for vision or language. Similarly, philosophical analysis remains crucial to clarify concepts and highlight potential pitfalls. Ongoing debates, such as whether consciousness requires a particular substrate (biological neurons vs. silicon) or whether it might be an emergent property of any sufficiently complex information system, will shape how we interpret advanced AI in the future. Some philosophers argue we might need entirely new paradigms (for instance, panpsychism or illusionism) to make sense of consciousness, which could radically affect how we attempt to implement or recognize it in machines.

In terms of practical milestones, a near-term goal is likely to be to develop empirical tests or benchmarks for consciousness-like attributes in AI. These would not claim to detect subjective experience directly (which may be impossible) but rather assess abilities associated with consciousness. For example, tests could evaluate an AI's degree of self-awareness, its flexibility in adapting global knowledge to novel problems, or its capacity for reporting on internal states. One proposed avenue is a sort of "AI consciousness spectrum" – a set of cognitive competencies (e.g., theory of mind, understanding of self versus others, temporal awareness of self) that could be measured. An AI that scores highly across many of these dimensions could be considered to have a higher degree of "AI consciousness" (in the access sense). Such frameworks would help move the discussion from abstract possibility to concrete progress: Researchers could then compete or collaborate on advancing AI along this spectrum, much as they do with benchmarks for intelligence.

Finally, ethical foresight must evolve in tandem with technical progress. As we inch closer to machines with human-like capabilities, even if still not conscious, we must continuously revisit our policies and perceptions. If an AI claims to be conscious or behaves in a way that is indistinguishable from a conscious agent, at what point do we err on the side of caution and consider granting it moral consideration? Some have suggested adopting a principle of "reasonable doubt": If we cannot be certain that a machine is not conscious, we should treat it gently – just in case. While we are not yet at that point, these discussions must begin now, so that society is not caught unprepared by the eventual emergence of machines with mind-like attributes.[85] Conversely, we also need to manage public expectations and prevent misconceptions. For example, consumers might assume a clever chatbot is a sentient companion when it is not, potentially leading to confusion or emotional harm. Clear communication about the capabilities and limitations of AI consciousness will thus remain the responsibility of experts in the field.

The present state of research suggests that artificial consciousness, in the rich sense of the term, is still more of a theoretical construct than a realized technology. Contemporary AI aligns with weak AI: Extraordinarily capable in narrow domains, but devoid of inner experience. However, the field is steadily laying the groundwork that may 1 day support at least the functional attributes of consciousness. By drawing on neuroscience to inform AI design (e.g., global workspaces and self-monitoring loops) and by deepening our theoretical understanding of consciousness (e.g., access vs. phenomenal, functional correlates of experience), we are inching toward the longstanding goal of a conscious machine. Whether that machine will feel anything, or whether we would recognize its feelings if it did, remains uncertain. What is clear is that this line of inquiry will continue to challenge our scientific ingenuity and our philosophical openness. The coming years will likely bring machines that blur the line between programmed behavior and adaptive, self-directed cognition even further. How we choose to interpret and interact with those machines will be a test of our wisdom, calling for a balanced approach that is at once scientifically rigorous, philosophically informed, and ethically attuned to both the possibilities and the limits of machine consciousness.

Each step forward forces us to refine our understanding of our own minds, as much as that of machines, reinforcing the notion that the quest for artificial consciousness is as much a mirror for humanity as it is a window into the future of technology. By rigorously exploring both the capabilities and the limitations of our creations – while keeping concepts like access and phenomenal consciousness in clear view – we can guide advancements in a responsible manner.[81,82] Ultimately, the effort to build or identify consciousness in an artificial entity will deepen our grasp of the nature of consciousness itself, and in doing so, it will bridge disciplines in unprecedented ways. The discussion presented here – synthesizing perspectives from cognitive neuroscience, computational modeling, and machine learning – underscores that achieving artificial consciousness is not simply an engineering challenge, but an interdisciplinary grand question – one that will likely occupy philosophers, scientists, and engineers for decades to come.

5. Conclusion

The question of whether machines can possess consciousness remains a central debate in AI. Strong AI envisions machines capable of genuine cognitive states and understanding, while weak AI suggests they only simulate thought processes. The creation of artificial consciousness represents a profound and unresolved challenge in AI research. Progress in understanding the mechanisms underlying human consciousness is essential for evaluating the feasibility of replicating these processes in machines. Although present AI systems lack true consciousness, advancements in neuroscience and machine learning offer promising avenues for further exploration in this interdisciplinary domain.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Seyed-Ali Sadegh-Zadeh
Writing – original draft: Seyed-Ali Sadegh-Zadeh
Writing – review & editing: Mahboobe Bahrami

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

Not applicable.

References

1. Crick F, Koch C. Consciousness and neuroscience. *Cereb Cortex*. 1998;8(2):97-107.

   doi: 10.1093/cercor/8.2.97

2. Ornstein RE. *The Psychology of Consciousness*. New York: Viking Press; 1972.

3. Baluška F, Reber A. Sentience and consciousness in single cells: How the first minds emerged in unicellular species. *Bioessays*. 2019;41(3):e1800229.

   doi: 10.1002/bies.201800229

4. Jourahmad Z, Habibabadi JM, Moein H, *et al*. *Machine Learning Techniques for Predicting the Short-Term Outcome of Resective Surgery in Lesional-Drug Resistance Epilepsy*. [arXiv Preprint]; 2023.

5. Velmans M. An introduction to the science of consciousness. In: Velmans M, editor. *The Science of Consciousness*. England, UK: Routledge; 2003. p. 13-34.

6. Nazari MJ, Shalbafan M, Ghannadi F, *et al*. A machine learning approach for differentiating bipolar disorder type II and borderline personality disorder using electroencephalography and cognitive abnormalities. *PLoS One*. 2024;19(6):e0303699.

   doi: 10.1371/journal.pone.0303699

7. Koch C. *The Quest for Consciousness*. Englewood, CO: Roberts and Company; 2004.

8. Sadegh-Zadeh SA, Bahrami M, Najafi A, Asgari-Ahi M, Campion R, Hajiyavand AM. Evaluation of COVID-19 pandemic on components of social and mental health using machine learning, analysing United States data in 2020. *Front Psychiatry*. 2022;13:933439.

   doi: 10.3389/fpsyt.2022.933439

9. Song C. Brain structural complexity and consciousness. *Philos Mind Sci*. 2021;2:6.

   doi: 10.33735/phimisci.2021.1020

10. Piolino P, Desgranges B, Eustache F. Episodic autobiographical memories over the course of time: Cognitive, neuropsychological and neuroimaging findings. *Neuropsychologia*. 2009;47(11):2314-2329.

    doi: 10.1016/j.neuropsychologia.2009.01.020

11. Piolino P, Desgranges B, Clarys D, *et al*. Autobiographical memory, autonoetic consciousness, and self-perspective in aging. *Psychol Aging*. 2006;21(3):510-525.

    doi: 10.1037/0882-7974.21.3.510

12. Mace JH, Hidalgo AM. Semantic-to-autobiographical memory priming affects involuntary autobiographical

memory production after a long delay. *Conscious Cogn*. 2022;104:103385.

doi: 10.1016/j.concog.2022.103385

13. Tyler C. The emergent dualism view of quantum physics and consciousness. *Cosmos Hist J Nat Soc Philos*. 2015;11(2):97-114.

14. Foulkes D, Fleisher S. Mental activity in relaxed wakefulness. *J Abnorm Psychol*. 1975;84(1):66-75.

doi: 10.1037/h0076162

15. Delacour J. An introduction to the biology of consciousness. *Neuropsychologia*. 1995;33(9):1061-1074.

doi: 10.1016/0028-3932(95)00033-Y

16. Nieminen JO, Gosseries O, Massimini M, *et al*. Consciousness and cortical responsiveness: A within-state study during non-rapid eye movement sleep. *Sci Rep.* 2016;6:30932.

doi: 10.1038/srep30932

17. Sadegh-Zadeh SA, Soleimani Mamalo A, Kavianpour K, *et al*. Artificial intelligence approaches for tinnitus diagnosis: Leveraging high-frequency audiometry data for enhanced clinical predictions. *Front Artif Intell*. 2024;7:1381455.

doi: 10.3389/frai.2024.1381455

18. Nir Y, Massimini M, Boly M, Tononi G. Sleep and consciousness. In: Laureys S, Tononi G, editors. *Neuroimaging of Consciousness*. Germany: Springer; 2013. p. 133-182.

doi: 10.1007/978-3-642-37579-6_7

19. Meijer DKF, Geesink HJH. Consciousness in the universe is scale invariant and implies an event horizon of the human brain. *Neuroquantology*. 2017;15(3):41-79.

doi: 10.14704/nq.2017.15.3.1079

20. Pereira A Jr. Triple-aspect monism: A conceptual framework for the science of human consciousness. In: Pereira A, Lehmann D, editors. *The Unity of Mind, Brain and World*. Cambridge: Cambridge University Press; 2013. p. 299-321.

doi: 10.1017/CBO9781139342134.015

21. Sadegh-Zadeh SA, Bagheri M. Harnessing the power of clinical data in dentistry: Importance and guidelines for dentists in AI modelling for enhanced patient care. *Open J Clin Med Images*. 2024;4(1):1188.

doi: 10.52768/2833-2725/1188

22. Jin SH, Chung CK. Messages from the brain connectivity regarding neural correlates of consciousness. *Exp Neurobiol*. 2012;21(3):113-119.

doi: 10.5607/en.2012.21.3.113

23. Oyebode F. *Sims' Symptoms in the Mind: Textbook of Descriptive Psychopathology*. 6th ed. Amsterdam: Elsevier Health Sciences; 2018.

24. Sadegh-Zadeh SA, Kambhampati C. All-or-none principle and weakness of Hodgkin-Huxley mathematical model. *Int J Math Comput Sci*. 2017;11:453-460.

25. Sadegh-Zadeh SA, Kambhampati C. Computational investigation of amyloid peptide channels in Alzheimer's disease. *J (Basel)*. 2018;2(1):1-14.

doi: 10.3390/j2010001

26. Schiff ND, Plum F. The role of arousal and "gating" systems in the neurology of impaired consciousness. *J Clin Neurophysiol*. 2000;17(5):438-452.

doi: 10.1097/00004691-200009000-00010

27. Margetis K, Korfias SI, Gatzonis S, *et al*. Intrathecal baclofen associated with improvement of consciousness disorders in spasticity patients. *Neuromodulation*. 2014;17(7):699-704.

doi: 10.1111/ner.12219

28. Mhuircheartaigh RN, Rosenorn-Lanng D, Wise R, Jbabdi S, Rogers R, Tracey I. Cortical and subcortical connectivity changes during decreasing levels of consciousness in humans: A functional magnetic resonance imaging study using propofol. *J Neurosci*. 2010;30(27):9095-9102.

doi: 10.1523/JNEUROSCI.5593-09.2010

29. Lane RD, Smith R. Levels of emotional awareness: Theory and measurement of a socio-emotional skill. *J Intell*. 2021;9(3):42.

doi: 10.3390/jintelligence9030042

30. Blumenfeld H. Brain mechanisms of conscious awareness: Detect, pulse, switch, and wave. *Neuroscientist*. 2021;29(5):9-18.

doi: 10.1177/10738584211049378

31. Menon V. Large-scale brain networks and psychopathology: A unifying triple network model. *Trends Cogn Sci*. 2011;15(10):483-506.

doi: 10.1016/j.tics.2011.08.003

32. Cavanna AE. The precuneus and consciousness. *CNS Spectr*. 2007;12(7):545-552.

doi: 10.1017/S1092852900015337

33. Gil R, Arroyo-Anlló EM, Ingrand P, *et al*. Self-consciousness and Alzheimer's disease. Acta Neurol Scand. 2001;104(5):296-300.

doi: 10.1034/j.1600-0404.2001.00334.x

34. Pins D, Ffytche D. The neural correlates of conscious vision. *Cereb Cortex*. 2003;13(5):461-474.

doi: 10.1093/cercor/13.5.461

35. Passingham RE, Bengtsson SL, Lau HC. Medial frontal cortex: From self-generated action to reflection on one's own

performance. *Trends Cogn Sci*. 2010;14(1):16-21.

doi: 10.1016/j.tics.2009.11.001

36. Heilman KM, Nadeau SE, Beversdorf DO. Creative innovation: Possible brain mechanisms. *Neurocase*. 2003;9(5):369-379.

doi: 10.1076/neur.9.5.369.16553

37. Laureys S, Gosseries O, Tononi G, editors. *The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology*. 2nd ed. United States: Academic Press; 2015.

38. Sadegh-Zadeh SA, Bagheri M, Saadat M. Decoding children dental health risks: A machine learning approach to identifying key influencing factors. *Front Artif Intell*. 2024;7:1392597.

doi: 10.3389/frai.2024.1392597

39. Chow HM, Horovitz SG, Carr WS, *et al*. Rhythmic alternating patterns of brain activity distinguish rapid eye movement sleep from other states of consciousness. *Proc Natl Acad Sci U S A*. 2013;110(25):10300-10305.

doi: 10.1073/pnas.1217691110

40. Blows WT. *The Biological Basis of Mental Health*. 4th ed. England, UK: Routledge; 2021.

41. Seth AK, Bayne T. Theories of consciousness. *Nat Rev Neurosci*. 2022;23(7):439-452.

doi: 10.1038/s41583-022-00580-4

42. Baars BJ. *A Cognitive Theory of Consciousness*. Cambridge: Cambridge University Press; 1993.

43. Sadegh-Zadeh SA, Nazari MJ, Aljamaeen M, Yazdani FS, Mousavi SY, Vahabi Z. Predictive models for Alzheimer's disease diagnosis and MCI identification: The use of cognitive scores and artificial intelligence algorithms. *NPG Neurol Psychiatr Geriatr*. 2024;24:194-211.

doi: 10.1016/j.npg.2024.02.002

44. Rosenthal DM. Consciousness and its function. *Neuropsychologia*. 2008;46(3):829-840.

doi: 10.1016/j.neuropsychologia.2007.11.012

45. Aleksander I. Artificial neuroconsciousness: An update. In: Mira J, Sánchez-Andrés JV, editors. *International Workshop on Artificial Neural Networks*. Germany: Springer; 1995. p. 566-583.

doi: 10.1007/BFb0020193

46. Franklin S, Baars BJ, Ramamurthy U, Ventura M. *The Role of Consciousness in Memory*; 2005. Available from: https://ccrg.cs.memphis.edu/assets/papers/wrcm-bmm.pdf [Last accessed on 2024 Oct 29].

47. Tulving E. Memory and consciousness. *Can Psychol*. 1985;26(1):1-12.

doi: 10.1037/h0080017

48. Kouider S, De Gardelle V, Sackur J, Dupoux E. How rich is consciousness? The partial awareness hypothesis. *Trends Cogn Sci*. 2010;14(7):301-307.

doi: 10.1016/j.tics.2010.04.006

49. Reggia JA. The rise of machine consciousness: Studying consciousness with computational models. *Neural Netw*. 2013;44:112-131.

doi: 10.1016/j.neunet.2013.02.010

50. Manzotti R. From artificial intelligence to artificial consciousness. In: Chella A, Manzotti R, editors. *Artificial Consciousness*. Exeter: Imprint Academic; 2007. p. 174-190.

51. Sadegh-Zadeh SA, Dastmard A, Kafshgarkolaei LM, *et al*. Machine learning modelling for compressive strength prediction of superplasticizer-based concrete. *Infrastructures*. 2023;8(2):21.

doi: 10.3390/infrastructures8020021

52. Cleeremans A, Jiménez L. Implicit learning and consciousness: A graded, dynamic perspective. In: Stadler M, Frensch P, editors. *Implicit Learning and Consciousness*. United Kingdom: Psychology Press; 2002. p. 1-40.

53. Chrisley R. Philosophical foundations of artificial consciousness. *Artif Intell Med*. 2008;44(2):119-137.

doi: 10.1016/j.artmed.2008.04.003

54. Vaneechoutte M. Experience, awareness and consciousness: Suggestions for definitions as offered by an evolutionary approach. *Found Sci*. 2000;5(4):429-456.

doi: 10.1023/A:1011371421112

55. Masakowski YR. Artificial intelligence and the future global security environment. In: Boulanin V, editor. *Artificial Intelligence and Global Security*. United Kingdom: Emerald Publishing Limited; 2020.

56. Wiedermann J, Van Leeuwen J. Finite state machines with feedback: An architecture supporting minimal machine consciousness. In: Cooper SB, Dawar A, Löwe B, editors. *Conference on Computability in Europe*. Germany: Springer; 2019. p. 286-297.

doi: 10.1007/978-3-030-29750-3_29

57. Reggia JA, Katz GE, Davis GP. Artificial conscious intelligence. *J Artif Intell Conscious*. 2020;7(1):95-107.

doi: 10.1142/S2705078520500059

58. Iantovics LB, Gligor A, Niazi MA, Biro AI, Szilagyi SM, Tokody D. Review of recent trends in measuring the computing systems intelligence. *Brain (Bacau)*. 2018;9(2):77-94.

doi: 10.18662/brain/15

59. Wang J, Wang C, Cai P, *et al*. Artificial sense technology:

Emulating and extending biological senses. *ACS Nano.* 2021;15(12):18671-18678.

doi: 10.1021/acsnano.1c07293

60. Kohli V, Tripathi U, Chamola V, Rout BK, Kanhere SS. A review on virtual reality and augmented reality use-cases of brain-computer interface based applications for smart cities. *Microprocess Microsyst*. 2022;88:104392.

doi: 10.1016/j.micpro.2022.104392

61. Henderson D, Horgan T, Potrč M, Strahovnik V. Chromatic illumination: Conscious intentionality without conscious representation. *ProtoSociology*. 2021;38:35-58.

doi: 10.5840/protosociology2021383

62. Raikov A. *Cognitive Semantics of Artificial Intelligence: A New Perspective*. Germany: Springer; 2021.

doi: 10.1007/978-3-030-73164-9

63. Searle JR. Minds, brains, and programs. In: *The Oxford Handbook of Philosophy of Cognitive Science*. Oxford: Oxford University Press; 2019.

64. Chella A, Pipitone A, Morin A, Racy F. Developing self-awareness in robots via inner speech. *Front Robot AI*. 2020;7:16.

doi: 10.3389/frobt.2020.00016

65. Jamieson GA. *Psychology of Consciousness: Theory, Research, and Practice*. United States: American Psychological Association; 2021.

66. Haikonen POA. On artificial intelligence and consciousness. *J Artif Intell Conscious*. 2020;7(1):73-82.

doi: 10.1142/S2705078520500035

67. Kulikov SB, Shirokova AV. Artificial intelligence, culture and education. *AI Soc*. 2021;36(1):305-318.

doi: 10.1007/s00146-020-01018-9

68. Wei L. Legal risk and criminal imputation of weak artificial intelligence. In: *IOP Conference Series: Materials Science and Engineering*. England: IOP Publishing; 2019. p. 062085.

doi: 10.1088/1757-899X/677/6/062085

69. Velik R. AI reloaded: Objectives, potentials, and challenges of the novel field of brain-like artificial intelligence. *Brain Broad Res Artif Intell Neurosci*. 2012;3(3):25-54.

doi: 10.18662/brain/3.3/20

70. Dong Y, Hou J, Zhang N, Zhang M. Research on how human intelligence, consciousness, and cognitive computing affect the development of artificial intelligence. *Complexity*. 2020;2020:1-10.

doi: 10.1155/2020/1681031

71. Salameh A. *Artificial Intelligence as a Commons-Opportunities and Challenges for Society*. Berlin: Springer; 2017.

72. Lieto A. *Cognitive Design for Artificial Minds*. England, UK: Routledge; 2021.

73. Heuveline V, Stiefel V. Artificial intelligence and algorithms: True progress or just digital alchemy? In: Jäger G, editor. *Intelligence-Theories and Applications*. Cham: Springer; 2022. p. 219-227.

doi: 10.1007/978-3-030-91177-5_15

74. Ng GW, Leung WC. Strong artificial intelligence and consciousness. *J Artif Intell Conscious*. 2020;7(1):63-72.

doi: 10.1142/S2705078520500023

75. Dubovikov K. *Managing Data Science: Effective Strategies to Manage Data Science Projects and Build a Sustainable Team*. United Kingdom: Packt Publishing Ltd.; 2019.

76. Searle JR. Minds, brains, and programs. *Behav Brain Sci*. 1980;3(3):417-424.

doi: 10.1017/S0140525X00005756

77. Ray A. *Compassionate Artificial Intelligence: Frameworks and Algorithms. Compassionate AI Lab*. United States: (An Imprint of Inner Light Publishers); 2018.

78. Levy D. The ethical treatment of artificially conscious robots. *Int J Soc Robot*. 2009;1(3):209-216.

doi: 10.1007/s12369-009-0022-6

79. Chatila R, Renaudo E, Andries M, *et al*. Toward self-aware robots. *Front Robot AI*. 2018;5:88.

doi: 10.3389/frobt.2018.00088

80. Kinouchi Y, Mackin KJ. A basic architecture of an autonomous adaptive system with conscious-like function for a humanoid robot. *Front Robot AI*. 2018;5:30.

doi: 10.3389/frobt.2018.00030

81. Hildt E. Artificial intelligence: Does consciousness matter? *Front Psychol*. 2019;10:1535.

doi: 10.3389/fpsyg.2019.01535

82. Block N. On a confusion about a function of consciousness. *Behav Brain Sci*. 1995;18(2):227-247.

doi: 10.1017/S0140525X0003842X

83. Dehaene S, Lau H, Kouider S. What is consciousness, and could machines have it? In: Dignum V, editor. *Robotics, AI, and Humanity*. Berlin: Springer; 2021. p. 43-56.

doi: 10.1007/978-3-030-51110-4_4

84. Gabriel M. Could a robot be conscious? Some lessons from philosophy. In: Dignum V, editor. *Robotics, AI, and Humanity*. Berlin: Springer; 2021. p. 57-68.

doi: 10.1007/978-3-030-51110-4_5

85. Gunkel DJ. *Robot Rights*. United States: MIT Press; 2018.

86. Darling K. Extending legal protection to social robots: The

effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In: Calo R, Froomkin AM, Kerr I, editors. *Robot Law*. United Kingdom: Edward Elgar Publishing; 2016.

87. Scheutz M. The inherent dangers of unidirectional emotional bonds between humans and social robots. In: Lin P, Abney K, Bekey GA, editors. *Robot Ethics: The Ethical and Social Implications of Robotics*. United States: MIT Press; 2011. p. 205-221.

88. Sadegh-Zadeh SA. *Computational Methods Toward Early Detection of Neuronal Deterioration [dissertation]*. United Kingdom: University of Hull; 2019.