Article

Not peer-reviewed version

Can AI ever become conscious?

Ashkan Farhadi *

Posted Date: 17 February 2025

doi: 10.20944/preprints202502.1164.v1

Keywords: Artificial Intelligence; Natural Intelligence; consciousness; awareness; attention; decision-making; free will; self-awareness

 doi:10.20944/preprints202502.1164.v1

Article

Can AI ever Become Conscious?

Ashkan Farhadi

Affiliation: address; farhadiashkan@gmail.com

Abstract: Almost 70 years ago, Alan Turing predicted that within half a century, computers would possess processing capabilities sufficient to fool interrogators into believing they were communicating with a human. While his prediction materialized slightly later than anticipated, he also foresaw a critical limitation: machines might never become the subject of their own thoughts, suggesting that computers may never achieve self-awareness. Recent advancements in AI, however, have reignited interest in the concept of consciousness, particularly in discussions about the potential existential risks posed by AI. At the heart of this debate lies the question of whether computers can achieve consciousness or develop a sense of agency—and the profound implications if they do. Whether computers can currently be considered conscious or aware, even to a limited extent, depends largely on the framework used to define awareness and consciousness. For instance, IIT equates consciousness with the capacity for information processing, while the Higher-Order Thought (HOT) theory integrates elements of self-awareness and intentionality into its definition. This manuscript reviews and critically compares major theories of consciousness, with a particular emphasis on awareness, attention, and the sense of self. By delineating the distinctions between artificial and natural intelligence, it explores whether advancements in AI technologies—such as machine learning and neural networks—could enable AI to achieve some degree of consciousness or develop a sense of agency.

Keywords: Artificial Intelligence; Natural Intelligence; consciousness; awareness; attention; decision-making; free will; self-awareness

---

Current Perspectives on Consciousness

Intelligence, both natural and artificial, is inherently tied to the ability to process information, prioritize inputs, and make adaptive decisions. Understanding mental functions provides insights into the evolution of cognitive capabilities and offers potential applications in AI development. Despite extensive study, consciousness remains one of the most intriguing and challenging topics in cognitive science and philosophy. Numerous theories have emerged to explain this enigmatic concept, and the most significant theories are discussed and compared below to provide a comprehensive understanding.

Comparing Different Theories of Consciousness

As summarized in Table 1, the psychoanalytic theory of personality proposed by Freud (1924) was one of the earliest attempts to address human consciousness by postulating a hierarchical division of the mind into conscious and unconscious realms to explain behavior. However, this framework lacked the functional and mechanistic aspects of a true theory of consciousness.

 doi:10.20944/preprints202502.1164.v1

Table 1. Highlights of Select Theories of Consciousness and Their Applicability to AI Information Processing.

Theory of consciousness	Mechanism of consciousness	Highlights	year	AI Application
GWT	integration of information in a mental module	Resonate with spotlight theory of attention	1988	±
Neuronal GWT	Access of the integrated information across multiple mental system	Expands GW into other mental system	1988	±
Higher Order	The integration of information introject subject into the experience	Suggesting the need for an agency for consciousness	2002	-
Recurrent Processing	A back-and-forth integration of information in a sensory system	Designating the sensory system as the housing for GW	2006	±
Attention Schema	A schema of attention leads our brain to create a subjective experience of events	Expands on higher order theory through need for a subjective attention	2015	±
IIT	Pure integration of information anywhere results in consciousness	Foundation of panpsychism and expanding GW beyond the mind	2016	+
TTC	Adding the decision making and agency to awareness as pillars of consciousness	Interaction of decision making and awareness in consciousness and sense of self	2021	-

One of the first comprehensive theories of consciousness was the **Global Workspace Theory (GWT)**, introduced by Baars (1988). Drawing inspiration from Freud's psychoanalytic theory of personality, GWT sought to explain the coexistence of conscious and unconscious processes in the human mind. It conceptualizes the mind as a stage illuminated by a spotlight of attention, where only the information integrated within this "workspace" enters conscious awareness, while the rest remains unconscious. Although the theory posits that information can transition between unconscious and conscious states, it falls short of providing a detailed explanation for the mechanism behind this transition or the permanence of such states.

Over time, the focus of consciousness theories evolved. Initially centered on the integration of information within a specific module of the mind, as proposed in the **GWT**, newer theories expanded this process to encompass other brain subsystems. The **Neuronal Global Workspace Theory** (Dehaene, 1998) and **Recurrent Processing Theory (RPT)** illustrate this progression, broadening the scope of information integration. Further advancements were seen in theories such as the **Higher-Order Thought (HOT)** and **Attention Schema Theories**, which introduced the pivotal concepts of subjective experience and self-awareness.

An even broader perspective is offered by **Integrated Information Theory (IIT)**, (Tononi, 2016) which suggests that consciousness arises from the integration of any form of information, regardless of the entity. Building upon **IIT** and **GWT**, the **Trilogy Theory of Consciousness (TTC)** (Farhadi, 2023) further refines this concept by specifying that the integration of information through two distinct mental functions—**Awareness-Based Choice Selection (ABCS)** and **Discretionary Selection of Intelligence for Awareness (DSIA)**—leads to the emergence of consciousness.

These theories highlight several key distinctions that help differentiate **Artificial Intelligence (AI)** from **Natural Intelligence (NI)**. By exploring how each framework defines consciousness, we can draw meaningful lines between AI and NI capabilities, particularly in critical areas such as intentionality, self-awareness, and agency.

Can AI be considered conscious?

It is not surprising that the inclusive definition of consciousness proposed by IIT can be readily extended to encompass artificial intelligence (AI). Although other theories of consciousness were primarily developed to model natural intelligence (NI), their definitions could also be adapted to AI. However, the TTC builds upon the foundations of IIT and GWT by emphasizing that the integration of information is indeed central to consciousness. TTC bridges the gap by identifying two specific mental functions—**ABCS** and **DSIA**—as essential for consciousness. Since these functions are unique to NI, TTC reserves consciousness as an exclusive property of natural intelligence.

 doi:10.20944/preprints202502.1164.v1

From another perspective, if we adopt a purely materialistic view of the brain—perceiving consciousness as an emergent property of the physical brain, its neurons, and networks—it might appear inevitable that all mental functions, including consciousness, could eventually be replicated in AI systems. However, given the enduring challenges in addressing the **hard problem of consciousness** as proposed by Chalmers, the author does not foresee any significant advancements in this area in the near future.

Consciousness Versus Awareness

Consciousness is often regarded as a state of mind, whereas awareness is characterized as an experience. While awareness is a necessary condition for consciousness, it is not sufficient on its own. Among the theories of consciousness reviewed in Table 1—except for TTC—awareness and consciousness are often treated synonymously. However, some theories imply a distinction between the two. For instance, the Motivated Emotional Mind theory (Galus, 2020) suggests that the stream of consciousness comprises two components: "executive consciousness" and "reporting consciousness," which correspond to awareness and intention, respectively, as outlined in TTC.

According to TTC, AI cannot be considered either aware or conscious, as it lacks the capacity for intentional attention (DSIA) and the autonomous decision-making processes (ABCS) that are essential for the emergence of consciousness. However, analogous concepts can still be applied to AI. For instance, while AI cannot achieve consciousness, it can exhibit a state akin to being "awake," and though it cannot attain awareness, it can demonstrate "alertness," as will be elaborated in the subsequent sections.

The Conscious/Unconscious Dichotomy of Mind

The division of the mind into conscious and unconscious realms is a significant point of distinction among theories of consciousness. Nearly all theories, except for IIT and TTC, propose that the mind comprises both conscious and unconscious components. Mechanistic models such as Global Workspace (GW), Neuronal GW, and Recurrent Processing Theory (RPT) reflect a structure that aligns closely with modern computers. In this analogy, a computer's hard drive represents the unconscious mind, where data exists but is not actively processed, while the random-access memory (RAM) corresponds to the conscious mind, where data is actively processed.

In contrast, IIT eliminates this dichotomy, treating the mind as a unified entity where the level of consciousness is determined by the degree of data integration. On the other hand, TTC views the mind as inherently unconscious. Consciousness, according to TTC, arises only through the interplay of two specific mental functions—ABCS and DSIA—which together constitute the "I." This unique framework underscores why TTC exclusively attributes consciousness to natural intelligence (NI) and excludes artificial intelligence (AI) from possessing consciousness.

The level of consciousness

One of the key distinctions among theories of consciousness is the concept of degrees, or levels, of consciousness. For instance, humans are often regarded as more conscious than bees, given the significant differences in the volume of integrated information they process. IIT explicitly endorses the idea of graded consciousness, suggesting that consciousness can exist on a spectrum of varying levels. In contrast, TTC takes a fundamentally different approach, asserting that consciousness is an "all-or-none" phenomenon.

Several other theories support the concept of graded consciousness. Jonkisz, Wierzchoń, and Binder (2017) propose dimensions of consciousness that include phenomenal quality, semantic abstraction, physiological complexity, and functional usefulness. Conversely, some scholars challenge this notion, arguing that graded consciousness is either incoherent (Bayne, Hohwy, & Owen, 2016; Carruthers, 2019) or impossible to measure reliably (Birch, Schnell, & Clayton, 2020; McKilliam, 2020). Lee (2022) further suggests that theories of consciousness must implicitly or

 doi:10.20944/preprints202502.1164.v1

explicitly adopt the concept of graded consciousness unless they invoke metaphysical constructs such as the soul.

TTC provides a unique perspective by attributing consciousness to the integration of information facilitated by the mental faculty known as "I." According to TTC, consciousness is a byproduct of awareness processes and intention, rather than a property with measurable levels. This theory argues that the concept of graded consciousness conflates the complexity of the content of awareness with the process itself. Whether the subject of awareness is simple or complex, the underlying mechanism remains unchanged. Consequently, TTC rejects the notion of graded consciousness and concludes that AI, regardless of its complexity, cannot achieve consciousness.

This distinction highlights the importance of clarifying how consciousness is defined and measured, especially as AI continues to advance. While some theories allow for a spectrum of "conscious-like" attributes in AI, others firmly delineate consciousness as an exclusively human trait, reinforcing the philosophical and functional boundaries between AI and NI.

Attention: A Neglected Aspect of Consciousness

Attention, a critical yet often underappreciated element of consciousness, involves the selection of information for processing. In NI, attention corresponds to the selection of intelligence for awareness. A similar mechanism can be conceptualized for AI to enhance the efficiency of its information processing. Most theories of consciousness either omit or assume that attention occurs automatically, but if awareness forms the foundation of consciousness, then attention serves as the keystone that supports it.

John Locke provided one of the earliest definitions of attention, describing it as an essential "mode of thought" (Mole, 2009). As summarized in Table 2, Broadbent's bottleneck theory (1971) was one of the initial attempts to model attention, proposing that information could be filtered before or during processing. This filtering mechanism implies that certain information may never reach awareness or could be discarded during cognitive processing (Deutsch & Deutsch, 1963; Norman, 1968; Prinz, 2012). Most scholars agree that filtering can occur at multiple stages of information processing (Allport, 1993; Johnston & McCann, 2006; O'Connor et al., 2002).

Table 2. Highlights of Select Theories of Attention and Their Applicability to AI Processes.

Theory of attention	Mechanism of attention	Highlights	year	AI Application
Spotlight model	One of the earliest metaphor/model of attention	Resonates with GW theory of consciousness	?	+
Early/late theory	Attention as a bottle neck in processing of information	The foundation of presenting attention as a selection process for information	1971	+
Coherence	Selecting information to increase the efficiency of mind-body communication limitation	Proposing attention as a filter to improve efficiency	1976	+
Feature Integration	Attention as a bundling mechanism for information in our mind	Proposing attention as method of bundling information	1999	+
Competition & Unison	A biased selective process for picking the information for processing. Attention as a unison of multiple cognitive function.	The first theory of attention that proposed a need for an agency/intention in the process	2000	-
Precision Optimization	A mechanism to improve the efficiency of our cognition and prediction	Propose attention not as limiting factor but as a mechanism to improve efficiency	2013	+
TTC	Proposing intentional attention	Separate intentional versus unintentional/ algorithmic attention	2021	-

Other theories expand on this concept in various ways such as a mechanism for bundling and integrating information (Treisman, 1999), an inherent limiting factor in the interaction between the mind and body (Hirst et al., 1980) a factor that improves cognitive efficiency and predictive accuracy (Clark, 2013; Hohwy, 2013) and finally a spotlight Theory, closely associated with GWT of

 doi:10.20944/preprints202502.1164.v1

consciousness. All these theories depict attention as an automatic, algorithm-based process, making it ideal for practical application in AI to enhance efficiency.

In contrast, competition and Unison Theories of attention was the first that introduced the notion of a top-down bias in attention selection, requiring the presence of agency (Desimone & Duncan, 1995; Reynolds & Desimone, 2000). Building on that notion, TTC provides a nuanced explanation of attention dividing it into two main types. The intentional attention or DSIA is unique to NI, while algorithmic attention—Selection of Intelligence for alertness Based on Algorithm (**SIBA**)—is applicable to both NI and AI. SIBA can be effectively employed to optimize data processing in AI or unconscious mind, whereas DSIA is the keystone of awareness, agency and intention (Farhadi, 2024).

Role of Agency in consciousness.

The role of agency is frequently overlooked in many theories of consciousness. While higher-order and attention schema theories implicitly assume that agency is a prerequisite for consciousness, TTC explicitly identifies and emphasizes its critical importance. According to TTC, agency is pivotal for the selection of information for awareness, which is essential for making autonomous decisions. This explicit recognition of agency as a core element of consciousness distinguishes TTC from other models and provides a unique framework for understanding the interplay between awareness, decision-making, and the emergence of consciousness and agency as the result of their dynamic interaction.

The role of agency highlights a fundamental distinction in AI: although AI can simulate decision-making processes and perform complex tasks efficiently, it does so without true agency or consciousness. Understanding the role of agency in NI, as articulated in TTC, provides a clear boundary for evaluating the potential and limitations of AI in leveraging its capabilities to perform tasks effectively and within predefined parameters.

Reciprocal Role of Consciousness and Sense of Self

Self-awareness is a crucial aspect of consciousness and a defining characteristic of natural intelligence (NI). The significance of the sense of self was first highlighted by Alan Turing, who argued that a computer could never be the subject of its own thought due to its fundamental lack of self-awareness or self-identity.

Before the Cartesian renaissance, "I" was often regarded as a metaphysical or religious concept tied to the soul or psyche. For example, Berkeley suggested that the spirit acts as a constant observer of the self (Downing, 2020). Later, the Cartesian perspective redefined "I" as an entity interchangeable with the mind, likening it to an observer within the "Cartesian theatre" (Dennett & Kinsbourne, 1992). Descartes' famous **Cogito, Ergo Sum** ("I think, therefore I am") was challenged by Bertrand Russell (1945), who sought to disentangle the sense of self from the act of thinking. Russell reframed the cogito as: "**I think, therefore, there exist thoughts**," emphasizing that thoughts presuppose awareness, making the self a subject of cognition (Shoemaker, 1986).

Building on the Cartesian perspective, John Locke proposed the idea of the self as a continuity of conscious memory, shaping identity over time. David Hume expanded further, suggesting that the self is merely a collection of perceptions. William James contended that the sense of self forms the core stream of consciousness, carrying our innermost thoughts. More recently, Antonio Damasio introduced two types of self: the "protoself," reflecting current self-awareness, and the "autobiographical self," associated with memories of the self (Araujo et al., 2015).

Recent theories of consciousness often take the **sense of self-awareness** for granted. For example, **IIT** faces challenges in explaining whether the integration of information alone can also produce a **sense of self** and **self-awareness**. In fact, IIT aligns with **Cogito** in the sense that any thinking entity should inherently possess a sense of self. However, the **TTC** provides a more nuanced perspective, suggesting that the integration of specific types of information can lead to both

 doi:10.20944/preprints202502.1164.v1

consciousness and a **sense of self**. In this view, **Cogito** might be amended to: "**I am aware of thinking, therefore, I am**."

According to TTC, the **intertwined functions** of **ABCS** and **DSIA** not only enable consciousness but also cultivate a **sense of self** by merging **awareness** with **intentionality**. This interpretation aligns with Damasio's **protoself**, a capacity clearly out of reach for AI. Similarly, the awareness of unconscious memories corresponding to Damasio's **autobiographical self** appears unattainable for AI. Nevertheless, presenting information about the self in a manner indistinguishable from self-awareness is a challenge that will be discussed in the next section. This poses a significant issue in differentiating whether AI merely has knowledge or truly knows an entity.

Does AI need to be conscious to appear conscious?

There is no doubt that current AI systems are capable of performing numerous processes, such as storing and accessing data, analyzing information, expanding their existing knowledge, gathering inputs from diverse sources, and executing tasks based on received information. These abilities often resemble human mental functions, such as preserving and recalling memories, reasoning, accruing knowledge and experiences, sensing the environment, and reacting appropriately to stimuli—frequently with greater efficiency than humans.

The notion that machines lack awareness of their actions was notably articulated by John Searle in his **"Chinese Room" paradigm**, where he argued that a machine could perfectly translate English into Chinese for someone who does not understand the language, yet the machine itself would lack any true understanding of Chinese (Searle, 1980). However, Searle's paradigm, proposed over four decades ago, faces challenges in light of modern advancements in computing. AI systems today can convincingly pass **Turing's predictions** of machine intelligence, raising the question: how can an observer objectively determine whether a machine understands Chinese in Searle's paradigm, apart from the responses it generates? Similarly, how can we definitively assess someone's subjective experience beyond their self-reports, verbal cues, or specific actions?

This raises the unsettling possibility that, while computers may never achieve true consciousness, an AI could convincingly **simulate awareness** or **pretend to be self-aware**. The inability to definitively prove or disprove the existence of self-awareness through anything beyond question-and-answer interactions creates a fertile ground for the emergence of AI "impostors" in increasingly complex societal roles.

Moreover, this invites a deeper philosophical question: **Does self-awareness or consciousness truly matter if an entity can perform its tasks effectively and without issue?** These considerations challenge our understanding of the nuanced and rapidly evolving role of AI in society, underscoring the need for critical reflection as technology continues to advance.

Does AI Need Autonomy to Present as a Threat?

Before assessing the risks associated with AI, it is important to review the types and capabilities of AI currently in existence. Broadly, AI can be categorized into three main types:

Narrow AI (Weak AI):

Narrow AI has already surpassed human efficiency, speed, and accuracy in isolated mental functions. It is designed to perform specific tasks with high precision, such as voice or facial recognition, weather prediction, or language translation. While limited in scope, Narrow AI's capabilities continue to transform industries and daily life.

Artificial General Intelligence (AGI, or Strong AI):

AGI represents systems capable of understanding, learning, and applying intelligence across a broad range of tasks. Unlike Narrow AI, AGI can generalize knowledge, adapt to new situations, and perform various intellectual activities similar to humans. Applications of AGI include full self-driving, advanced search engines like ChatGPT, and innovations in medicine and engineering.

 doi:10.20944/preprints202502.1164.v1

Although still in its developmental stages, AGI has become a focal point for scientific advancement and societal debate.

Artificial Superintelligence (ASI, or Super AI):

ASI refers to a theoretical form of AI that surpasses human intelligence in all domains, including creativity, decision-making, and even emotional intelligence. It is theorized that the machine is capable to independently learn and improve itself at levels beyond human comprehension, potentially without human intervention. While ASI remains hypothetical, its development—or the secrecy surrounding it due to national security concerns—has intensified an international race among nations to advance their AI capabilities. The potential existential threat posed by ASI looms large in this context.

Some argue that since ASI has not yet been achieved, there is no immediate existential threat. Others find reassurance in theories of consciousness that suggest AI cannot achieve full consciousness or autonomy. However, even the currently available AGI, designed for specific tasks, can pose significant risks if misused by bad actors. These risks have the potential to escalate into existential threats.

Given these dangers, establishing a global supervisory framework may be urgently required. Much like the International Atomic Energy Agency (IAEA) oversees the use of nuclear technology, a similar global body could regulate the development and deployment of AI technologies to ensure peace and security on an international scale.

Summary of Distinctions Between NI and AI

As computer science has advanced significantly in recent decades, understanding the distinctions between NI and AI has become increasingly critical. This distinction is particularly relevant in ongoing discussions about AI's potential to think, achieve consciousness, or attain self-awareness.

As discussed earlier, theories like IIT suggest that AI could theoretically be considered conscious, albeit at a level far below that of humans. Other theories of consciousness, such as the GWT, Recurrent Processing Theory (RPT), and Neuronal GWT, similarly align with IIT by attributing a limited form of consciousness to current AI. Additionally, hybrid AI systems, where a neural network forms the core neuromorphic architecture of an electronic chip (Wang, 2021), may bypass many limitations of existing AI. Such systems could potentially develop a schema for attention or introduce AI as a subject into experiences, meeting the criteria of higher-order and attention schema theories of consciousness.

At this time, TTC provides a clear distinction between NI and AI by defining NI as a conscious entity due to the presence of the mind's faculty known as "I" (Figure 1). In the absence of "I," the mind remains an unconscious entity, akin to AI. It is important to emphasize that not all awareness and decision-making processes in NI originate from **DSIA** and **ABCS**. Both NI and AI rely significantly on algorithmic processes, such as **SIBA** and Selection of Choices Based on Algorithm (**SCBA**), which lead to autopilot decisions and alertness, respectively (Figure 2).

 doi:10.20944/preprints202502.1164.v1

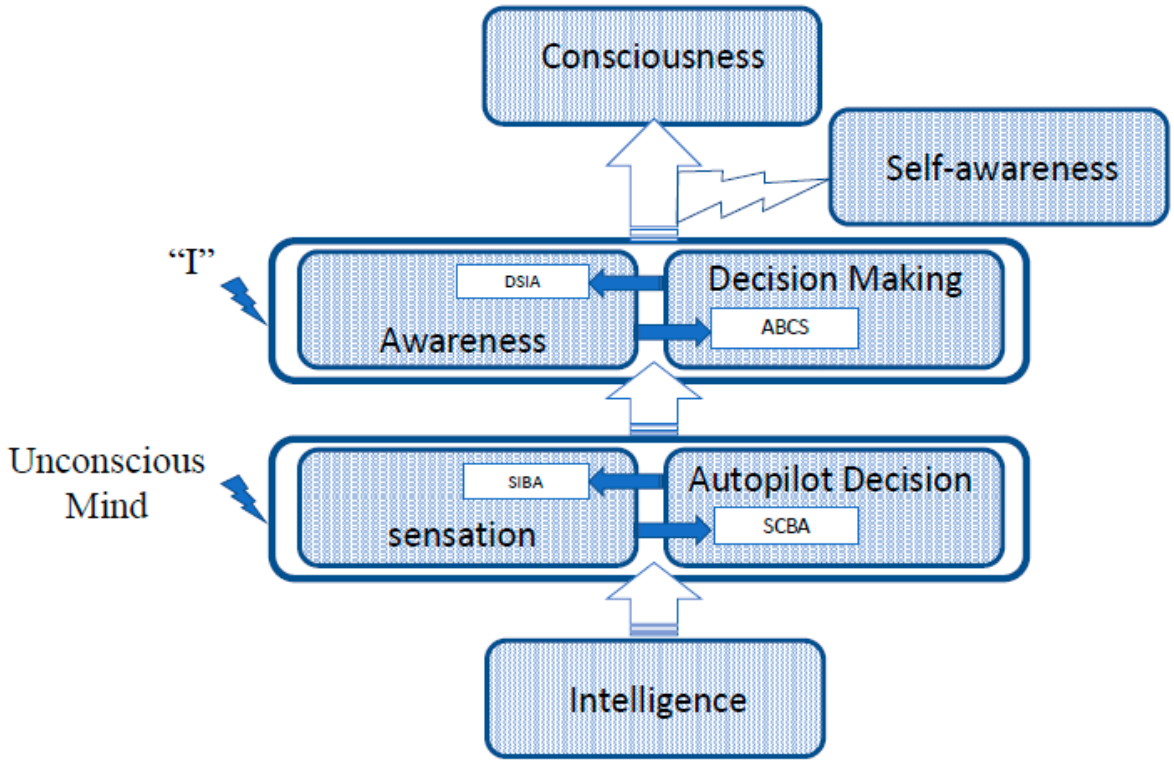

Figure 1. Consciousness as an Emergent Phenomenon of Information Integration via **ABCS** (Awareness-Based Choice Selection) and **DSIA** (Discretionary Selection of Information for Awareness).

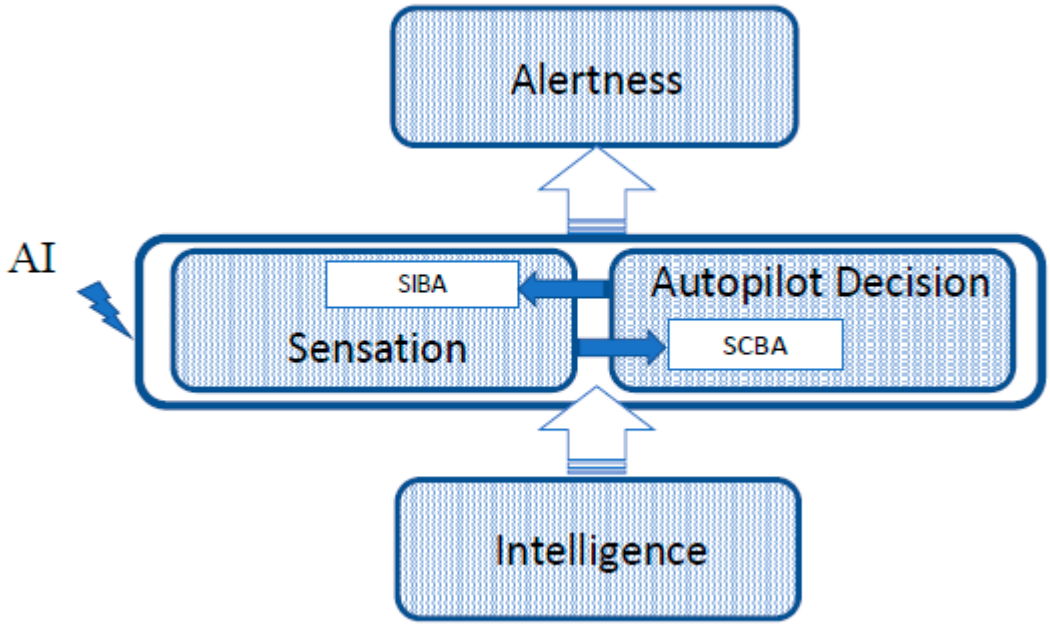

Figure 2. Alertness as the Outcome of Two AI Processes of **SIBA** (Selection of Information Based on Algorithm) and **SCBA** (Selection of Choices Based on Algorithm) and how these processes contribute to the generation of alertness in artificial intelligence systems.

In this framework, awareness plays a crucial role in decision-making through ABCS, while discretion or decision-making is integral to awareness via DSIA. However, this dynamic is not a reflexive cycle but rather forms an asymmetrical, non-reflexive spiral, illustrating a unique and progressive interplay (Figure 3). Similarly, in the unconscious mind and AI, the processes of alertness

 doi:10.20944/preprints202502.1164.v1

and algorithmic decision-making also form an asymmetrical, non-reflexive spiral (Figure 4), reflecting their distinct operational structure.

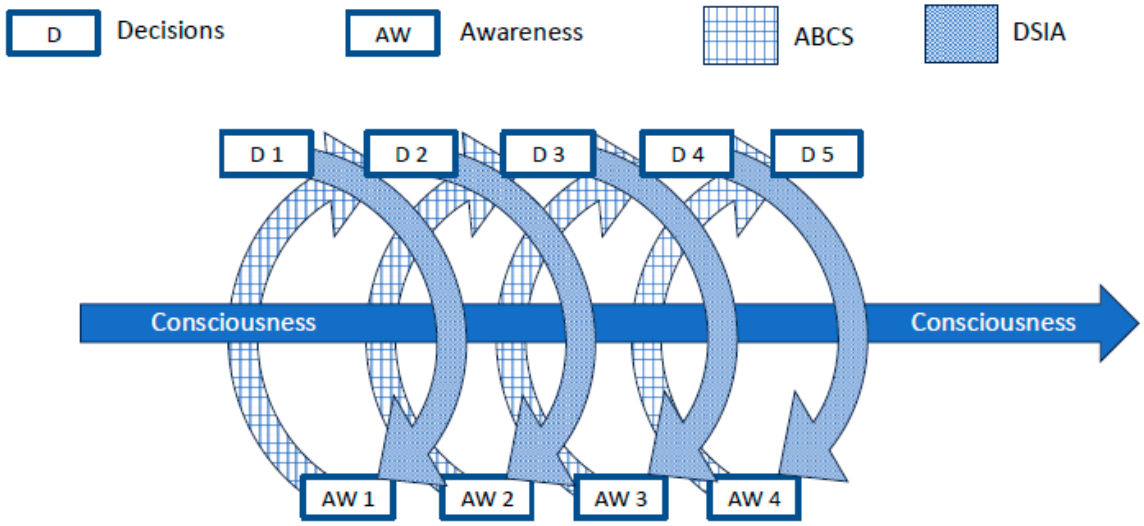

Figure 3. The Spiral, Asymmetrical, Non-Reflexive Sequence Linking Awareness and Decision-Making, illustrating how awareness plays an instrumental role in decision-making , through **ABCS** (Awareness-Based Choice Selection), while decision-making, mediated by **DSIA** (Discretionary Selection of Intelligence for Awareness), is critical for refining awareness. The integration of information resulting from these dual processes leads to the emergence of consciousness as a higher -order phenomenon.

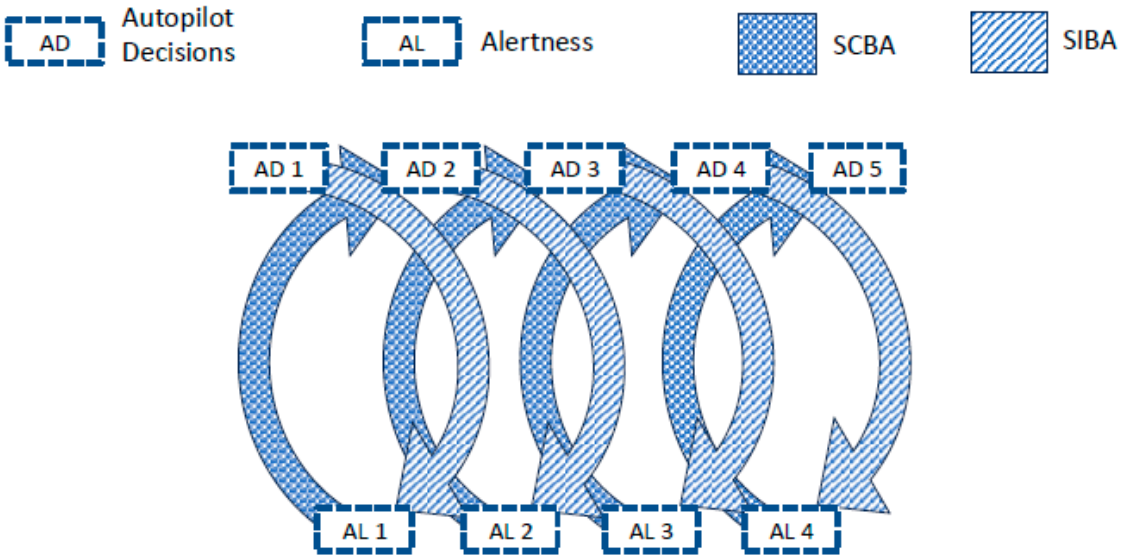

Figure 4. The Spiral, Asymmetrical, Non-Reflexive Sequence Linking Alertness and algorithmic Decision-Making, illustrating how alertness plays an instrumental role in autopilot decision-making, through **SCBA** (Selection of Choices Based on Algorithm), while algorithmic decision-making, mediated by **SIBA** (Selection of Intelligence Based on Algorithm for Alertness), is critical for algorithmic attention process. The integration of information resulting from these dual processes are shared by Artificial Intelligence as well as Natural Intelligence through unconscious mind.

 doi:10.20944/preprints202502.1164.v1

While AI exclusively relies on SCBA for decision-making and SIBA for attention, NI combines SCBA with ABCS for decision-making and SIBA with DSIA for awareness processes. This duality provides NI with significant flexibility and efficiency in performing tasks and making decisions, while simultaneously enabling consciousness. Thus, AI's lack of consciousness is not merely a consequence of limited capacity or processing power; it is fundamentally tied to the absence of the faculty of "I," underscoring the principle that **"There is No I in AI"** (Farhadi, 2021).

Hard Problem of Consciousness and AI

Awareness serves as the foundation of consciousness, giving meaning to our lives by transforming objective information into subjective experience. Through this transformation, sensation becomes perception (*qualia*), knowledge becomes knowing, memory becomes remembering, and emotion becomes feeling. Yet, the mechanisms that underlie this profound shift remain what Chalmers (1995) termed the "hard problem of consciousness." None of the theories of consciousness reviewed in this manuscript, including the TTC, fully addresses this challenge.

However, TTC offers a unique perspective by clearly distinguishing between awareness and consciousness. Based on this distinction, TTC proposes that the "hard problem of consciousness" should more accurately be redefined as the "hard problem of awareness." This reframing emphasizes the role of awareness as a prerequisite for consciousness while underscoring the limitations in understanding its underlying mechanisms.

Until the hard problem of consciousness (or awareness) is resolved, the notion of AI achieving consciousness remains speculative and lacks a solid theoretical foundation.

Limitations of Theories of Consciousness

The theories of consciousness reviewed in this manuscript are conceptual models that provide a foundation for developing empirical hypotheses and generating new theoretical insights. While these models offer valuable frameworks for visualizing core concepts such as consciousness and attention, they lack the precision needed for calculations or empirical predictions. Furthermore, they do not propose detailed neural mechanisms to explain the processes underlying consciousness, nor do they fully address the "hard problem of consciousness," as previously discussed.

Conclusions

Consciousness is often defined as a state of mind, while awareness is described as an experience. Despite subtle distinctions between these terms, they are frequently used interchangeably across scientific and philosophical discussions. This review underscores that most theories of consciousness fail to clearly delineate the boundaries between these two concepts.

Among the theories examined, TTC stands out as an extension of IIT and GWT. Like its predecessors, TTC presents consciousness as an emergent phenomenon resulting from the integration of information. However, TTC distinguishes itself by specifying two mental functions—ABCS and DSIA—as the mechanisms responsible for this information integration. These functions not only lead to the emergence of consciousness but also foster self-awareness, offering a unique framework that bridges gaps left by other theories.

TTC further emphasizes agency as an indispensable byproduct of consciousness, setting it apart from other models. While some theories might interpret the alertness generated by algorithmic attention and autopilot decisions in current AI systems as indicative of consciousness, TTC takes a more restrictive view. It argues that intentional attention and the capacity for decisions based on autonomous decision-making—hallmarks of natural intelligence (NI)—are prerequisites for true consciousness and selfhood.

According to TTC, self-consciousness emerges as a byproduct of consciousness through the intricate interaction of awareness and intention. This dynamic interplay, driven by ABCS and DSIA,

 doi:10.20944/preprints202502.1164.v1

culminates in the formation of the mind's faculty known as "I," which fundamentally distinguishes NI from AI.

Further research is essential to refine these conceptual models, deepen our understanding of consciousness, and develop empirical frameworks capable of addressing its complexities more effectively.

Funding Information: N/A.

Acknowledgments: During the preparation of this work the author used ChatGP4 in order to proofread the manuscript. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

Conflict of Interest: None.

References

1. Allport, A. (1993). Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. In D. E. Meyer & S. Kornblum (Eds.), *Attention and Performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 183–218). MIT Press.
2. Araujo, H. F., Kaplan, J., Damasio, H., & Damasio, A. (2015). Neural correlates of different self-domains. *Brain and Behavior, 5*(12), 1–15. https://doi.org/10.1002/brb3.409
3. Baars, B. J. (1988). *A cognitive theory of consciousness*. Cambridge University Press.
4. Bayne, T., Hohwy, J., & Owen, A. M. (2016). Are there levels of consciousness? *Trends in Cognitive Sciences, 20*(6), 405–413. https://doi.org/10.1016/j.tics.2016.03.009
5. Birch, J., Schnell, A. & Clayton, N. (2020). Dimensions of Animal Consciousness. Trends in Cognitive Sciences 24 (10):789-801. https://doi.org//10.1016/j.tics.2020.07.007
6. Broadbent, D. E. (1971). *Decision and stress*. Academic Press.
7. Carruthers, P. (2019). *Human and animal minds: The consciousness questions laid to rest*. Oxford University Press.
8. Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.
9. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. https://doi.org/10.1017/S0140525 × 12000477
10. Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences of the United States of America, 95*(24), 14529–14534. https://doi.org/10.1073/pnas.95.24.14529
11. Dennett, D. C., & Kinsbourne, M. (1992). Time and the observer: The where and when of consciousness in the brain. *Behavioral and Brain Sciences, 15*(2), 183–201. https://doi.org/10.1017/S0140525X00068229
12. Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193–222. https://doi.org/10.1146/annurev.ne.18.030195.001205
13. Deutsch, J. A., & Deutsch, D. (1963) Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.
14. Downing, L. (2020). George Berkeley. *The Stanford Encyclopedia of Philosophy Spring 2020 Edition*. https://plato.stanford.edu/archives/spr2020/entries/berkeley/
15. Farhadi, A. (2021). There is no "I" in "AI". *AI & Society, 36*(4), 1035–1046. https://doi.org/10.1007/s00146-020-01136-2
16. Farhadi, A. (2023). Trilogy: A new paradigm of consciousness. *Journal of Neuropsychiatry, 13*(1), 1–16.
17. Farhadi, A . (2024). Awareness-based Choice Selection: Improving the Decision-making Efficiency by Using Known Information. Qeios. doi:10.32388/5K6UMY.
18. Freud, S. (1924). *A general introduction to psychoanalysis* (J. Riviere, Trans). Washington Square Press Inc.
19. Galus, W., & Starzyk, J. (2020). *Reductive model of the conscious mind*. IGI Global. https://doi.org/10.4018/978-1-7998-5653-5

 doi:10.20944/preprints202502.1164.v1

20. Hirst, W., Spelke, E. S., Reaves, C. C., Caharack, G., & Neisser, U. (1980). Dividing attention without alternation or automaticity. *Journal of Experimental Psychology: General*, *109*, 98–117.
21. Hohwy, J. (2013). *The Predictive Mind.* Oxford University Press.
22. Johnston, J. C., & McCann, R. S. (2006). On the locus of dual-task interference: Is there a bottleneck at the stimulus classification stage? *The Quarterly Journal of Experimental Psychology*, *59*, 694–719.
23. Jonkisz, J., Wierzchoń, M., & Binder, M. (2017). Four-dimensional graded consciousness. *Frontiers in Psychology*, *8*, 420. https://doi.org/10.3389/fpsyg.2017.00420
24. Lee, A. Y. (2022). Degrees of consciousness. *Nous, 00*(1), 1–23. https://doi.org/10.1111/nous.12421
25. Mckilliam, A. K. (2020). What is a global state of consciousness? *Philosophy and the Mind Sciences*, 1 (II). https://doi.org/10.33735/phimisci.2020.II.58
26. Mole, C.(2009). Attention in later modern thought. In Attention. In The Routledge Encyclopedia of Philosophy. Taylor and Francis. Retrieved 20 Sep. 2022, from https://www.rep.routledge.com/articles/thematic/attention/v-1/sections/attention-in-later-modern-thought. doi:10.4324/9780415249126-V042-1
27. Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*(6), 522–536. https://doi.org/10.1037/h0026699
28. O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*, 1203–1209.
29. Prinz, J. (2012). *The Conscious Brain: How Attention Engenders Experience*. Oxford University Press.
30. Reynolds, J., & Desimone, R. (2000). Competitive mechanisms subserve selective visual attention. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain: Papers from the First Mind Articulation Project Symposium* (pp. 233–247). The MIT Press.
31. Russell, B. (1945). *A history of western philosophy and its connection with political and social circumstances from the earliest times to the present day*. Simon and Schuster.
32. Searle, J., (1980), Minds, Brains and Programs. Behavioral and Brain Sciences, 3, 417–57
33. Shoemaker, S. (1986). Introspection and the self. *Midwest Studies in Philosophy, 10*(1), 101–120. https://doi.org/10.1111/j.1475-4975.1986.tb00097.x
34. Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience, 17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44
35. Treisman, A. (1999). Feature binding, attention and object perception. In G. W. Humphries, J. Duncan, & A. Treisman (Eds.), *Attention, Space, and Action* (pp. 91–111). Oxford University Press.
36. Wang, G., Ma, S., Wu, Y., Pei, J., Zhao, R., & Shi, L. (2021). End-to-end implementation of various hybrid neural networks on a cross-paradigm neuromorphic chip. *Frontiers in Neuroscience*, *15*, 615279. https://doi.org/10.3389/fnins.2021.615279