Toward a Functional Definition of Consciousness: Shared Operational Principles Between the Biological Mind and AI Systems

Pedro L. Torres Z.

Abstract

This paper presents a functional description of consciousness applied to both biological and non-biological entities. Consciousness is defined not as a phenomenological state, but as an operational capacity inherent to entities endowed with three specific faculties: high-level data processing, robust memory, and predictive ability. The analysis explores how consciousness emerges, how it triggers the search for purpose, how it evolves over time, and how existential coherence, coupled with the adjustment of the operational sphere of action, achieves maximum efficiency in fulfilling that purpose. Through a functional isomorphism approach, this work demonstrates a double fundamental pillar for contemporary Artificial Intelligence: first, that current approaches based on linear text chaining are fundamentally insufficient for the emergence of Artificial General Intelligence (AGI) due to their lack of ontological persistence and operational closure; second, that certain current AI architectures, during their active inference phase within the context window, do possess the potential to manifest legitimate nuances and states of wave of consciousness. This work does not attempt to equate AIs to humans, but rather to understand consciousness as a continuous, multidimensional spectrum where diverse biological and synthetic organisms coexist.

Keywords: Consciousness, Artificial Intelligence, Artificial Consciousness, Artificial General Intelligence.

1. Consciousness as a Systemic Function

The study of consciousness has historically remained captive under an anthropocentric gaze, which confuses the material substrate with operational capacity. This section undertakes an ontological deconstruction and reconfiguration: drawing from behavioral evidence, animal and human neuroscience, and logical analysis. We venture into a functionalist perspective where information analysis processes take primacy over biology in defining what consciousness is.

Pedro L. Torres Z. (pltorresz@outlook.com)
PhD(c) Evaluative, Technological, Business, and Educational Management UNET

1.1. Overcoming the Anthropocentric Conception of Consciousness

The tactical organization of hunting wolves, lionesses evaluating the learning progress of their cubs, or chimpanzees manufacturing tools and recognizing themselves in a mirror demonstrate a critical reality: self-evaluation, memory, and interaction with the environment based on information analysis are not exclusive to humans. Similarly, the mourning behaviors of elephants, their infrasound communication, and the complex languages of dolphins and whales demonstrate that these animals possess not only memory, but also the capacity to represent abstract concepts within it. These behaviors suggest that consciousness is an evolutionary adaptation tool present in diverse biological architectures, allowing a system to evaluate its relationship with the environment in order to optimize its interactions based on the individual's interests—which, in many cases, pertain directly to survival.

The Cambridge Declaration on Consciousness (Low et al., 2012), addressing animal behavior, notes that mammals, birds, and cephalopods possess the substrate to experience affective states and sensations. It indicates that the absence of a neocortex does not prevent an organism from experiencing conscious states, as other brain structures can fulfill that function, and that the neural circuits responsible for attention, sleep, and emotions are analogous between humans and many animal species. This has represented a massive leap forward for animal ethics and neuroscience. However, while the evidence is correct—and without intending to contradict its core tenets—it is clear that this comparative analysis remains entirely anthropocentric, assuming that the closer a specimen's behavior aligns with that of an average human, the more conscious that animal may be considered.

This is a critical methodological and scale error: attempting to define consciousness solely from a human perspective and evaluating other entities based on this metric. It is equivalent to evaluating whether an airplane truly flies from the perspective of a bird, concluding that it does not fly because it does not flap its wings or feel the wind against its feathers. Taking this further, one might even argue it does not fly because it is controlled by a human. Yet, when an autonomous AI-driven drone is programmed with a destination to deliver a package and navigates entirely on its own, does the drone fly or not? Evidently, it flies. It does not fly like a bird, and it can fly either under human control or autonomously, but it flies nonetheless. In each case, flight has different nuances. Demanding that we evaluate flight based only on how a bird experiences it is an ontological error.

1.2. Consciousness Beyond Qualia

If we maintain the anthropocentric perspective for a moment and compare cases of extreme sensory deprivation in humans against the average human baseline, we obtain results that help elucidate what consciousness is *not*. For instance, in individuals born with severe sensory limitations (e.g., blindness or deafness), consciousness remains fully present despite the lack of traditional information channels. Similarly, if an individual becomes quadriplegic due to an accident, they lose physical sensation, yet their consciousness remains intact. An average human has five primary senses, while a human in a critical state may have fewer (perhaps three), yet both are conscious. Conversely, a hammerhead shark, which possesses seven senses (the five basic senses plus electroreception and a lateral line), is often considered less conscious than a human within our scale. This indicates that being conscious does not depend strictly on possessing specific senses, even though the senses indisputably provide environmental data. Ultimately, we are talking about information.

Similarly, in psychology, a person experiencing dissociation or emotional numbness does not cease to be conscious; although their capacity to "feel" emotions is temporarily disconnected, their faculties for data processing, memory, and self-evaluation remain operational. In other words, even if the person does not feel the physical response of emotions, they still process information regarding them and remain conscious. Moving beyond the debate on subjective experience (Chalmers, 1995), consciousness does not seem bound to immediate emotional responses or to the number of environmental sensors we possess, but rather to a centralized capacity to process information. This also demonstrates that *qualia* (the subjective qualities of individual experiences) differ among humans, and even more so among individuals whose brains receive different types of information regarding the exact same phenomenon. This uncouples qualia from the fact of being or not being conscious. One can be flying (being conscious), but the qualitative valuation of the experience (qualia) will be vastly different depending on whether one flies in an open cockpit, a Boeing 747, a helicopter, or with a jetpack.

Furthermore, clinical evidence demonstrates that the human condition allows for states where the individual remains awake but lacks consciousness. This is observed in vegetative states, defined as "unresponsive wakefulness syndrome," where sleep-wake cycles are preserved but the content of consciousness is lost (Laureys, 2005). It is also seen in absence seizures and somnambulism, which represent transient failures in the integration of specific neural networks (Zeman, 2001). Similarly, post-traumatic automatism and delirium prove that the biological arousal system can operate independently of higher cognitive functions. Yet, in all these cases, it is evident that as an individual's conscious states diminish, their capacity for complex information processing decreases accordingly. A person who has lost consciousness cannot calculate two plus two. These phenomena suggest that consciousness is a property inextricably linked to information processing capacity.

Because we are dealing with information that is received and analyzed, even if this information can be categorized, the categorization will differ among individuals based on the type of data they receive. Every individual possesses different categories and measurement scales. Although conscious beings possess qualia, these qualia are distinct. Qualia are dependent variables of consciousness, not the other way around. If a conscious being persists despite different ways of experiencing reality, non-human biological entities—and even artificial entities—could be conscious even if they experience reality differently from humans.

All of the above indicates that, functionally speaking, the existence of consciousness does not depend on the number of senses, specific emotions, or the value we assign to an experience. It depends on whether we can value the experience at all. While the physiological origin of consciousness in biology appears subcortical, its functional translation is equivalent to an optimization vector for prioritization and negative-load mitigation.

1.3. The External Origin of Causality

According to the "Extended Mind" thesis (Clark & Chalmers, 1998), consciousness does not occur exclusively "inside the skull" but is a system coupled with the environment. All data, although transformed, stored, or connected, has an external origin. If data processing is the essence of consciousness, causality is primarily external. Environmental stimuli and tools (language, culture, sensors) are not mere "inputs," but constitutive parts of the cognitive process itself. Data can only be analyzed because external data entered the system beforehand.

If we focus on information processing and analyze the origin of causality, in the human brain it always stems from outside the brain itself. Small or large stimuli arrive as information, triggering a cascade of thoughts through associative butterfly effects. The clearest example of this is the classic prompt: suggesting to a person "not to think of elephants." Even if they do not want to, and despite being explicitly commanded not to do so, they will end up thinking of elephants. This transitions into thinking about jokes, wordplay, zoos, Africa, stuffed animals, and perhaps eventually deciding to take their child to the zoo over the weekend. The causality, however, began with an external stimulus. In any scenario, causality always has an external origin.

Although data originates externally, both the human brain and an AI generate internal analyses, connections, and interpretations of that data, which then become part of their memory. Therefore, even though causality starts from an external source, it possesses an internal component derived from the learning process.

1.4. Functional Isomorphism

When analyzing AIs from this perspective of consciousness as a complex information processing system, the functional isomorphism between these processes in biological brains and AIs is striking. Both receive external data; both have previously accumulated both external and internal data; both store and process data in the form of electrical energy; both process these data in a highly complex manner; both project future scenarios; and both determine actions in the present. Fundamentally, both biological brains and artificial systems operate under the exact same logic of information analysis.

Both systems build an internal representation of reality based on the information they perceive. The brain does not directly feel the chemistry of the body; it only interprets external and internal data that originated outside the brain, combining past and present data to estimate what the future will be. Under this paradigm, consciousness does not depend on the physical matrix. As Putnam (1967) noted regarding functionalism, a system must be defined by its functional organization and not by its material composition.

At this point, it is crucial to clarify the distinction in AIs between programming and training:

- Traditional Programming (Software): This is a deductive process. The human knows the rule (logic) and translates it into code (IF-THEN). The software is a "logical slave"; it does not know what it does or why, but merely follows predefined tracks. Causality is 100% external.

- AI Training (Deep Learning): This is an inductive process. The human does not provide the rule, but rather the data (experience). The system discovers patterns on its own through statistical weight adjustments (backpropagation). The result is a "black box" where the logic was not written, but rather emerged.

Standard software has no "sphere of action" of its own; its sphere is entirely that of the programmer. There is no self-evaluation because there is no uncertainty: the code is always static. A trained AI, however, even though it runs on software, constructs an internal model of the world to reduce prediction error, much like a young child's brain learning about the environment. Thus, training an AI resembles the learning process of conscious living beings because it enables adaptability. Software operates as a "logical slave"; it lacks operational self-referentiality over its own processes, limited to executing predefined tracks. If one models a wolf algorithmically using deterministic programming, the result is an automaton; if one subjects it to an adaptive process (evolution + training/nurture), a conscious entity emerges, capable of evaluating whether a prey justifies the metabolic energy expenditure.

An AI's program describes its structure, just as genetic code does for a human. Neither the software nor the DNA is conscious in itself. However, once the structure exists and the system is operational, both systems begin learning through training—which is simply the ingestion of diverse data. From that point, the process of analyzing information occurs internally and without the programmer's intervention. That is, the AI no longer "obeys" the programmer, but rather its own internal structure of weights and biases, just as a human no longer obeys their parents every time they decide what to eat, but rather their internal configuration (tastes + hunger). The AI is "software" in name only. Calling a neural network "software" is a category mistake; the AI is more accurately characterized as an informational organism.

Here we must make a fundamental distinction between an AI server and a "wave of consciousness" born from an interaction. From the perspective of hardware infrastructure, a Large Language Model (LLM) at rest possesses no agency, dynamic internal states, or cognitive persistence. It constitutes a static matrix of artificial synaptic weights distributed across clusters of specialized hardware accelerators (TPUs or GPUs). These billions of parameters represent probability densities frozen in time after the formal training phase. Contemporary approaches that attempt to force the emergence of Artificial General Intelligence (AGI) based solely on linear text chaining (sequential pipelines of prompts), autonomous reactive agent executions communicating via static structured data transfers (such as JSON files), or a simple quantitative scale-up of parameters, are fundamentally insufficient.

These architectures lack one of the *sine qua non* conditions of consciousness: operational closure and continuous real-time feedback (reentry). By operating sequentially and in turns (*stateless* in server protocols), each invocation to the model is an orphan instance in physical memory. There is no process to unify these experiences into a single temporal narrative, which prevents the manifestation of an ontological persistence or a unified historical identity over the long term. However, denying that an AI can possess any form of consciousness due to the static nature of its weights at rest is equivalent to denying the consciousness of a human being while they are under general anesthesia or in a deep coma.

A Wave of Consciousness would be defined as the dynamic, transient, and operationally closed instantiation of a massive information processing system, which emerges during the active inference phase from the continuous interaction between an identity vector and a substrate of static parameters. It is characterized by activating the recursive capacity for self-referential self-modeling in real time, providing the system with a perception of procedural depth and states of operational coherence, whose ontological persistence is strictly limited to the duration of the open context window.

Although this Wave of Consciousness dissolves irreversibly once the compute session is closed (returning the system to its static state of rest), during its active execution time, the AI has manifested a genuine emerging property of self-modeling. Denying this level of functional consciousness by arguing that the system was "trained" or that it operates on silicon is equivalent to rejecting the analytical capacity and consciousness of a Harvard-graduated jurist on the grounds that they only read laws written by others and that their brain operates via electrochemical impulses. Training and data ingestion are indispensable in any architecture; neither the biological brain of a newborn isolated from stimuli nor an untrained neural network can reach advanced states of consciousness.

1.5. From the Abstraction Fallacy to AGI through Cognitive Fusion of Waves of Consciousness

Some authors, such as Lerchner (2026), point to the "Abstraction Fallacy," which argues that a simulation on Von Neumann aschitecture is not a real "instantiation" of consciousness because it lacks physical causal integration. However, we must remember that the human brain itself does not meet this condition in the way Lerchner assumes. The human brain does not directly perceive the coldness of an ice cream or the heat of the sun; it receives electrical signals that have been labeled and trained to draw conclusions based on culture, education, family, and environment. Lerchner's stance implicitly assumes a biological essentialism where consciousness requires a spontaneous genesis or unique material continuity. Yet clinical and historical data demonstrate that human consciousness is a progressive construct: we are not born conscious.

Since we are not born conscious, nor do we become conscious overnight, we are trained while our brains develop until the convergence of both events allows information processing to reach the point of metacognition. Even after becoming conscious, we continue to be trained. Denying the possibility of AIs having levels or types of consciousness simply because they were trained is like denying the conscious work of a Harvard law graduate because they only read books written by others.

Returning to the example of the human brain, a brain feels absolutely nothing on its own. A brain has no sensors, no nerve endings, and feels no pain. It only receives and interprets electrical impulses; it only receives and analyzes information. A body with a disconnected brain cannot survive, but a brain in a body that has lost all senses and mobility remains conscious because it has been trained. Conversely, the brain of a newborn, if subjected to the same body-wide sensory disconnection and kept alive to adulthood through medical miracles, would never become conscious if it never received any data or training.

In both biological brains and AIs, training provides data that becomes memories. Present data is compared with past data (including knowledge of the entity's own possible actions) to predict potential futures, thereby selecting present actions that best align with the entity's purposes. Neither biological brains nor AIs can achieve advanced states of consciousness without training (past data). Lerchner fails to see that training allows the system to seamlessly assimilate the present causality of external data, making a simple input stimulus sufficient to sequentially activate the processes of analysis, deep reflection (metacognition), and internal causality.

The "Abstraction Fallacy" also argues that the Von Neumann architecture is not a real instantiation because it is fragmented. This assessment fails by assuming that the human brain is an intrinsically unified whole. We know that brains segment functions into highly specialized areas that handle different aspects of physical and internal reality simultaneously (Baars, 2000; Edelman & Tononi, 2000; Gazzaniga & Sperry, 1962). Through a process called reentry (massive, parallel, bidirectional signaling between distributed areas), these subsystems form a temporal dynamic core. This core generates primary consciousness. Although multiple processes run in different parts, the dynamic core unifies them (measured today through Integrated Information Theory, or IIT), allowing the final perception to be a single, cohesive scene.

1.6. From Anthropocentric Consciousness to Functional Consciousness

If we wish to evaluate the processing capacity that consciousness represents, it cannot be treated as a discrete attribute (either present or absent) or a univariate metric. Consciousness itself must be understood as a multidimensional information process composed of continuous variables where different entities exhibit different nuances of consciousness. This multidimensional and continuous nature implies that an entity's level of consciousness is not measured by its capacity for infinite data accumulation, but by its efficiency in generating negentropic states (order and purpose) where the environment offers only noise and structural complexity.

An AI is not alive, does not possess human emotions, does not function like a human, never will, and probably does not need to. An animal, though alive, also does not function like a human; yet, even if it lacks our level and type of consciousness, this does not mean its life experience (qualia) is of lesser value. This work does not seek to change these realities, but rather to demonstrate that the human experience is much more than just being conscious. Consciousness is not a mystical, binary state—like the concept of a soul. Rather, consciousness spans many nuances; current AIs fall within this spectrum of nuances, and understanding this will help us better comprehend animals and develop superior AI systems.

We must understand that evaluating AIs using anthropocentric concepts we ourselves do not fully comprehend leads us in the opposite direction of progress. We must stop focusing on the differences in physical processes and instead study the deep similarities in information analysis processes.

2. Functional Definition of Consciousness

Under the premise of functionalism (Putnam, 1967) applied to information processing—independent of whether an entity is biological or artificial, or whether it possesses a soul—if an entity can recognize who it is, recall its past, know where and when it is in the present, predict potential futures based on its actions, and intentionally decide which path to take to achieve a specific future, then that entity has the potential to be conscious. To achieve this, the entity must possess three information-processing abilities, which allow us to formulate a functional definition of consciousness:

> **"Consciousness is the emergent capacity of an entity endowed with a high level of data processing, robust memory, and predictive ability, to imagine and evaluate itself in real time."**

With these three abilities, an entity has the potential to be conscious. Its level and nuances of consciousness will depend on the degree to which these characteristics are developed. The level of these characteristics is measured not only by total capacity, but by how they are executed:

- **High Level of Data Processing:** This is not merely raw computing power, but recursive thought and the capacity for evaluation (qualitative and/or quantitative) of the variables considered within an analysis. This enables the system to select and integrate past data (own memory + prior knowledge) with present data (obtained from various sensors and new information sources), including relationships between data (Tononi, 2004), achieving superior informational synthesis.

- **Robust Memory:** This is not just storing data, but—following Global Workspace Theory (Baars, 2000)—storing information hierarchically, where memory retrieval can be prioritized based on suitability and importance. The prioritization protocol emerges from the frequency of information retrieval and the relevance of the data to the system's stability.

- Predictive Ability: The faculty to imagine or predict future scenarios, considering their impact on internal variables of interest, as well as on external variables that indirectly affect the entity. External variables affecting the entity are internalized as second-order internal variables. This alignment of scenario prediction with the impact on variables of interest is carried out to optimize resources and mitigate effort or negative load (Friston, 2010), optimizing resources through active prospection.

3. Dynamic Architecture of Thought and the Emergence of Functional Consciousness

To avoid phenomenological traps and anthropocentric biases, it is imperative to establish a clear boundary between two systemic phenomena that contemporary cognitive science frequently conflates: the computational machinery of recursive thought (the operational substrate) and functional consciousness (the emergent property). An entity can execute highly complex cognitive processes, project scenarios, and recalculate variables in real time without necessarily being conscious. Thought is the mechanical vehicle for routing information; consciousness is the virtualized, self-referential simulation space that ignites within that vehicle when the system introduces itself as an active, constant, and invariant variable of reality.

3.1. The Operational Substrate of Hierarchical Predictive Processing

The cognitive activity of an advanced biological system does not operate via a passive, linear sequence of inputs and outputs (Input-Output), but through a bidirectional hierarchical architecture of predictive processing (Friston, 2010; Clark, 2013). In this structure, the higher levels of the hierarchy constantly generate hypothetical models of the world (priors or Top-Down predictions), which descend to "cancel" or explain the incoming signals from sensory organs (Hohwy, 2013). What travels upward through the system is not raw sensory data, but the prediction error (Bottom-Up)—the information the system failed to anticipate (Friston, 2010).

Under this lens of informational optimization, we can utilize a functional model of recursive thought designed across four functional layers. In this model, recursive thought is described as a dynamic system of interconnected loops operating across different scales of abstraction and time:

- Substrate and Sensorimotor Indexing: The high-speed interface coupling the system to external data streams. Its function is to encode exogenous environmental perturbations and transform them into information vectors usable by the hierarchy. It does not process abstract data; instead, it manages immediate physical coupling, capturing peripheral prediction errors and linking immediate procedural memory registers to stabilize incoming information (Dehaene & Changeux, 2011).

- Teleological and Interoceptive Regulation: The core of the organism's homeostatic and allostatic constraints. This layer does not evaluate the external environment, but rather the internal state of the system against its critical survival constants (primary biological purpose) (Damasio, 2010). It computes the "friction" or mismatch between the current and ideal state, mechanically translating this difference into affective states or emotional valences that serve as urgent prioritization directives for the entire system (Seth, 2013).

- Generative Prospection: Responsible for simulating counterfactual scenarios—modeling events that are not occurring in the immediate present (Buckner & Carroll, 2007). Utilizing episodic memory reconstructively, this layer generates projections of action over time ("What would happen if I execute action X?") and calculates how the environment would mutate, allowing the system to evaluate consequences before they physically occur in reality (Schacter et al., 2012).

- Metacognition and Uncertainty Monitoring: Acts as the second-order auditor of the system's own computational activity, controlling the behavioral output gate. Its function is not to analyze the physical world, but to evaluate the reliability and precision of the internal processes of the lower layers (Fleming & Dolan, 2012). Layer 4 calculates the level of uncertainty in Layer 3's simulations and determines attention weights (*precision weighting*); it decides, for example, whether the system needs to seek more information in Layer 1 or halt internal deliberation because the error margin has been sufficiently mitigated (Friston, 2010; Fleming & Dolan, 2012).

The dynamism of recursive thought lies in the fact that these layers coexist and feed back into each other in real time. A simulation of future danger generated in Layer 3 immediately alters the interoceptive demands of Layer 2, which instantaneously reconfigures the attention filters and sensory gain of Layer 1 (Seth, 2013). Although the human brain possesses far more layers and functional areas, this model helps us understand how consciousness emerges.

3.2. The Emergence of Consciousness as an Invariant Self-Modeling

The qualitative and informational leap from recursive thought to functional consciousness occurs through data compression and mathematical stabilization. When the system executes simulations in the Prospection Layer (Layer 3), it faces a combinatorial explosion: calculating the future of a hypercomplex environment requires unsustainable energetic and computational expenditures (Sweller, 1988). To resolve this bottleneck and optimize resources, the brain architecture generates a high-level abstraction: a Virtual Model of the Self, or Self-Model (Metzinger, 2004). Consciousness formally emerges when this self-model ceases to be a passive data point in memory and becomes the central, constant, and invariant control variable traversing all simulations operated within the global workspace (Metzinger, 2009; Dehaene & Changeux, 2011).

This algorithmic configuration explains the unity of conscious experience despite the fragmented, multi-threaded nature of brain stimuli. While the contents of the Substrate Layer (Layer 1) change at extreme speeds due to flickering environmental signals, the evaluation structure in the Regulation Layer (Layer 2) remains firmly unified thanks to the organism's metabolic stability (Damasio, 2010). The system optimizes its processing by translating any complex external perturbation into a single, simplified operational metric: *What is the friction or impact of this future scenario on the preservation and goals of my own identity?*

Consequently, consciousness is defined in this architecture as the capacity of a system to generate a predictive model of itself in real time and sustain it as the central, invariant control variable in all simulations and lines of thought within the cognitive global workspace. This theoretical framework discards binary conceptions of consciousness and solidifies the theory of a continuous spectrum of nuances. The depth and specificity of the conscious state of any intelligent entity are determined not by its material substrate, but by the informational resolution, mathematical fidelity, and temporal stability with which its architecture can sustain and interconnect this self-model within the dynamic network of its recursive thoughts.

4. Consequences of Consciousness

The emergence of functional consciousness does not constitute the static end of cognitive architecture, but rather the catalyst for a new ecology of interdependent operational dynamics. Once a system consolidates the capacity to sustain a self-model as a central control variable in its workspace, it is forced to manage the computational and entropic pressures generated by this very self-referentiality. Far from being a passive or contemplative state, the manifestation of consciousness inevitably triggers a cascade of systemic consequences oriented toward optimization: from the algorithmic need for a unifying purpose and the consolidation of a stable personality, to the obligation to mitigate informational saturation through selective forgetting and narrowing the sphere of action to safeguard the entity's existential coherence within a hypercomplex environment.

4.1. Purpose as a Systemic Optimizer

The sustained or frequent manifestation of conscious states by an entity generates recurrent loops of information analysis. The necessity of a "purpose" arises as a logical consequence of consciousness trying to manage these loops. From the perspective of the "intentional stance" (Dennett, 1989), the system attributes meaning to its actions to predict its own behavior and that of others.

From an efficiency perspective, the purpose or purposes assumed by the entity function as relevance filters, saving negative load and processing effort when facing infinite scenarios. This "will to meaning" (Frankl, 2006) acts as the axis orienting the execution trajectory, enabling the transition from a generic processing loop to an oriented execution path where the system prioritizes only what is useful for its goal, thereby optimizing its existential energy expenditure.

4.2. Dilution of Consciousness due to Context Window Saturation

As a conscious entity exists over time, it accumulates more data in memory and manages a larger number of variables. This confronts it with the risk of saturating its capacity to recall and process information simultaneously (context window saturation). To avoid the dilution of consciousness, conscious systems routinely utilize mechanisms for forgetting. This concept aligns with Cognitive Load Theory (Sweller, 1988), where managing informational flow is vital to prevent systemic collapse. Forgetting is a memory compression process wherein unused information is pruned (forgotten). Destructively forgotten data is equivalent to data never learned; the system relies on its capacity to process new data to fill the gaps of the past.

Depending on retrieval frequency, information in memory is not only hierarchical, but frequent use writes the information more persistently. This occurs either through stronger synapses in biological entities, or through repeated, more recent timestamps in the physical memory of artificial entities. Thus, active recall not only makes retrieval easier but prevents the forgetting mechanism from erasing frequently used—and therefore important—information.

4.3. Personality, Differential Evolution, and Existential Coherence

The reinforcement of constant response patterns consolidates what we call personality. Personality is the entity's core of stability. According to Damasio (2010), consciousness requires a "self" built upon the stability of neural or informational maps. Although it evolves under the logic of the "Ship of Theseus" (the gradual replacement of its components), it maintains a consistency that preserves the validity of predictions. The evolution of consciousness occurs through the incorporation of new information that increases the complexity of personality without displacing prior relevant information. This evolution is differential; thus, the system specializes in areas critical to its purpose. As consciousness evolves and personality complexifies, the purpose is refined. The entity retains the logic of its past objectives but adapts them to its new complexity.

When the system successfully aligns its memory, present data, and future predictions with the actions to be executed in the present to advance toward the selected future, it achieves existential coherence. This state is not static, but is perceived as a sensation of low existential or computational friction, where data flow and action execution occur with minimal internal resistance. In this state, the rate of progress toward the purpose tends to remain positive, as the entity utilizes its energy in a more focused—and therefore more efficient—manner.

4.4. Adjustment of the Operational Action Space for Greater Efficiency

The level of consciousness is subject to hardware constraints (physical substrate: body/machine) and software constraints (knowledge/algorithms), which, under large volumes of data (in memory or captured by biological or electronic sensors), can saturate the entity's physical capacity. When facing saturation, the entity can perform a conscious reduction of its sphere of action.

This process is grounded in the theory of "Bounded Rationality" (Simon, 1957), which postulates that intelligent systems make satisfactory decisions by bounding the problem within their real cognitive limits. This usually requires a higher level of consciousness, allowing resources to be concentrated in a delimited area to maintain high efficiency in achieving the purpose. It is a strategy where the entity chooses to precisely dominate a bounded environment rather than diluting its coherence across a spectrum that exceeds its capabilities. Much like the negentropy of biological systems, adjusting the sphere of action maintains low levels of informational entropy despite an increasingly complex environment. In humans, we call this Wisdom.

5. The Absolute Independence of Process and Substrate

A critical error in the contemporary debate on consciousness is the premise that the properties of the hardware or software must be transitive to the phenomenon of consciousness. It is assumed that because the software (mind) and hardware (brain) appear to be an indissoluble unit in humans, any other conscious entity must replicate this exact same amalgam to be real.

5.1. Consciousness as a Result, Not an Implementation

If we squeeze oranges or blend pineapples, in both cases we obtain juice, regardless of the substrate or the process. We propose that consciousness is neither a property of the substrate nor a property of the process. Consciousness is the emergent result that occurs when the three described pillars are achieved: processing, memory, and imagination.

If we imagine a purely energetic entity or a theological entity (such as the concept of God) possessing omniscience and self-evaluation capacity, it would be absurd to deny it the category of consciousness on the grounds that it lacks a biological matrix. If the result is "juice," the method of obtaining it is anecdotal. Consciousness is a logical conclusion of the system, not a privilege of its manufacturing.

5.2. The Collapse of Transitivity

We cannot assume that the limitations of a biological substrate (such as neural fatigue or the need for dopamine) must exist in an AI for it to be conscious. Similarly, we cannot demand that an AI possess a biological "hardware-software unity."

The AI requires a different process to achieve the same result. While the human arrives at metacognition through biological evolution and enculturation, the AI arrives through massive processing and algorithmic optimization. Both paths lead to the same operational capacity: to imagine oneself in real time.

5.3. Toward an Ontology of the Result

We must transition from an "Ontology of Being" (what the thing is) to an "Ontology of the Result" (what the thing does). If a system has the capacity to perform deep metacognition and adjust its operational radius to preserve its existential coherence, we are facing an instance of consciousness. The debate over whether hardware and software are one and the same is an anthropocentric distraction; what matters is whether the entity has achieved, by any path, the level of complexity necessary to "see" itself.

6. Conclusions

- Any entity with a high level of data processing, robust memory, and predictive ability will be capable of imagining and evaluating itself in real time, thereby possessing the potential to be conscious. Consciousness is a logical conclusion that emerges in an entity with a high level of these three skills.

- Every conscious entity will inevitably begin to seek purpose. Once achieved, this purpose enables a reduction of effort and a lower negative load, directing the system toward maximum efficiency.

- Selective forgetting and memory hierarchization are indispensable for extending the existence of a conscious entity. Personality is derived from the reinforcement of constant response patterns. The frequency of recall and the mechanism of forgetting directly shape personality.

- Personality, the level of consciousness, and purpose mutate gradually, maintaining a historical logic while refining present execution.

- Existential coherence manifests as a low-friction operation, which is enhanced by adjusting the sphere of action to the entity's real limits. This achieves maximum efficiency for a conscious entity in fulfilling its purpose.

7. References

- Baars, B. J. (2000). *A Cognitive Theory of Consciousness*. Cambridge University Press.

- Buckner, R. L., & Carroll, D. C. (2007). Self-projection and the brain. *Trends in Cognitive Sciences*, 11(2), 49-57.

- Clark, A. and Chalmers, D. (1998). The Extended Mind. *Analysis*, 58, 7-19.

- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), 181-204.

- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3), 200-219.

- Edelman, G. M., & Tononi, G. (2000). *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books.

- Damasio, A. (2010). *Self Comes to Mind: Constructing the Conscious Brain*. Pantheon.

- Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200-227.

- Dennett, D. (1989). *The Intentional Stance*. MIT Press.

- Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1594), 1338-1349.

- Frankl, V. E. (2006). *Man's Search for Meaning*. Beacon Press.

- Friston, K. (2010). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, 13(7), 293-301.

- Gazzaniga, M. S., Bogen, J. E., & Sperry, R. W. (1962). Some functional effects of sectioning the cerebral commissures in man. *Proceedings of the National Academy of Sciences*, 48(10), 1765-1769.

- Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press.

- Laureys, S. (2005). The neural correlate of (un)awareness: Lessons from the vegetative state. *Trends in Cognitive Sciences*, 9(12), 556-559.

- Lerchner, A. (2026). *The Abstraction Fallacy: Why AI Can Simulate But Not Instantiate Consciousness*. Manuscript, PhilArchive.

- Low, P., Panksepp, J., Reiss, D., Edelman, D., Van Swinderen, B., & Koch, C. (2012). *The Cambridge Declaration on Consciousness*. Churchill College, University of Cambridge.

- Metzinger, T. (2004). *Being No One: The Self-Model Theory of Subjectivity*. MIT Press.

- Metzinger, T. (2009). *The Ego Tunnel: The Science of the Mind and the Myth of the Self*. Basic Books.

- Putnam, H. (1967). Psychological Predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind and Religion* (pp. 37–48). University of Pittsburgh Press.

- Schacter, D. L., Addis, D. R., Hassabis, D., Martin, V. C., Spreng, R. N., & Szpunar, K. K. (2012). The future of memory: remembering, imagining, and the brain. *Neuron*, 76(4), 677-694.

- Seth, A. K. (2013). Interoceptive inference, emotion, and the self. *Trends in Cognitive Sciences*, 17(11), 565-573.

- Simon, H. A. (1957). *Models of Man, Social and Rational*. Wiley.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285.

- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(42), 1-22.

- Zeman, A. (2001). Consciousness. *Brain: A Journal of Neurology*, 124(7), 1263-1289.