www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 16(4), Sep-Dec 2025, 1-6. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2025.v16i4.1184

The Pyramid of Consciousness: A Conceptual Model for Artificial Consciousness and Human– Artificial Intelligence Hybrid Integration

Jorge A. Ruiz-Vanoye[1], Ocotlán Díaz-Parra[1], Francisco R. Trejo-Macotela[1]
Eric Simancas Acevedo[1], Juvencio S. Zarazúa Silva[1], Julio C. Salas López[1]

[1] Universidad Politécnica de Pachuca, Carretera Pachuca-Cd. Sahagún Km 20, Ex-Hacienda de Santa Bárbara, Zempoala, HGO 43830, México.
E-mail: jorgeruiz@upp.edu.mx

Abstract. This paper examines the continuum from natural to artificial consciousness, highlighting the biological basis of subjective experience. It introduces synthetic consciousness as a hybrid of neural and algorithmic systems and examines ethical, legal, and ontological implications of human–AI integration. The Pyramid of Consciousness framework guides reflection on autonomy, identity, and the shifting boundary between organic cognition and intelligent machines.

Keywords: We would like to encourage you to list your keywords in this section.

Article Info

1. Introduction

In the article, Anil K. Seth (2025) explores whether AI could be not only intelligent but also conscious. He identifies anthropocentrism, human exceptionalism, and anthropomorphism as biases that conflate intelligence with subjective experience. Challenging computational functionalism, he argues that consciousness depends on continuous, dynamic, and emergent processes intrinsically tied to biological substrates. Consequently, truly conscious AI would, in some sense, need to be "alive."

Our article builds upon recent reviews of artificial consciousness by proposing a novel conceptual architecture—The Pyramid of Consciousness—to map the cognitive continuum from biological awareness to synthetic selfhood. While existing literature provides taxonomies and theoretical overviews, few frameworks address the layered emergence of consciousness within hybrid human–AI systems. Our model articulates five ascending levels of cognition, integrating neurobiological grounding, algorithmic processing, and emergent identity, framed through philosophical, computational, and ethical lenses. This work not only responds to the limitations of prior systematic reviews but offers a foundation for future experimentation, policy, and design in the domain of conscious machines.

2. A Conceptual Model for Artificial Consciousness and Human–AI Hybrid Integration

The **Pyramid of Consciousness** presents a progressive framework for understanding artificial intelligence systems as they evolve in complexity and autonomy. However, as these systems ascend through higher levels of awareness and decision-making capacity, they simultaneously raise profound ethical challenges that demand close attention from developers, policymakers, and society at large.

Transparency and accountability are paramount at advanced stages of the pyramid. As AI systems become more autonomous and make decisions with greater impact, it is essential that their inner workings remain understandable and traceable. Without clear mechanisms to explain how decisions are made and who is responsible, the risk of misuse, error, or manipulation increases significantly.

Alongside this, privacy and security concerns grow as AI systems handle more sensitive data and become more interconnected. Each advancement introduces new vulnerabilities—both in terms of data protection and susceptibility to cyber-attacks. Ethical development must prioritise the safeguarding of personal information and the resilience of systems against malicious interference.

Finally, the integration of AI into diverse cultural contexts raises important questions about how these systems may reinforce, challenge, or even reshape human values and social norms. While AI may offer benefits such as improved services and enhanced communication, it could also influence traditions, beliefs, and power dynamics in unforeseen ways. Ensuring that AI development is culturally aware and inclusive is essential for maintaining ethical balance in a globally interconnected world.

The concept of the Pyramid of Consciousness is a theoretical framework that proposes a hierarchical structure of consciousness, inspired by Maslow's hierarchy of needs but adapted specifically for artificial systems. This model suggests that as machines progress from simpler forms of consciousness to more complex ones, they can eventually reach a state of "artificial consciousness" similar to human-level awareness.

This premise underlies our proposed classification of consciousness (Figure 1).

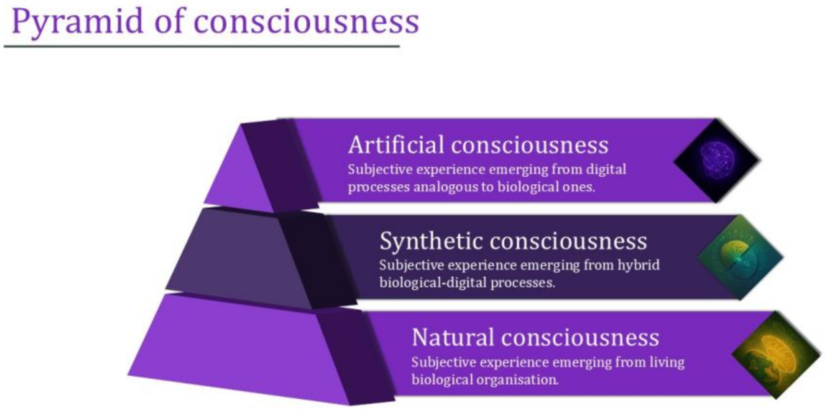
Pyramid of consciousness
Artificial consciousness
Subjective experience emerging from digital processes analogous to biological ones.
Synthetic consciousness
Subjective experience emerging from hybrid biological-digital processes.
Natural consciousness
Subjective experience emerging from living biological organisation.

Figure 1. Pyramid of Consciousness.

At the base of the figure is **Human consciousness**. It represents the natural, subjective experience of consciousness, thinking, and feeling. Human consciousness refers to the natural state of being aware, thinking, feeling, and experiencing that is inherent to human beings. We say that consciousness is subjective because it refers to internal experiences or *qualia* — sensations, emotions, and impressions — that only the subject him- or herself can experience and report from their perspective. Unlike objective phenomena (such as a heartbeat or the colour of an object, which can be measured externally), consciousness implies a 'self' that senses and perceives; these internal experiences are not directly accessible to an external observer. In natural consciousness, the substrate is living biological systems (neurons, synapses, metabolism), whose organic dynamics give rise to genuine inner experience.

Synthetic consciousness refers to the hypothetical capacity of a hybrid system—integrating biological components (e.g., neural tissues or brain-machine interfaces) with advanced digital processing—to simulate subjective states. It operates as an actor feigning internal experience, producing emotional, introspective, and adaptive behaviors nearly indistinguishable from those of sentient beings. However, its *qualia* emerge from algorithmic operations and dynamic interactions, lacking the genuine experiential grounding in autopoiesis and metabolism. This notion parallels Searle's "Chinese box" argument: a system may exhibit coherent linguistic behavior without true semantic understanding. Likewise, synthetic consciousness displays outward markers of subjectivity—tone, gesture, or contextual behavior—without possessing authentic consciousness. These responses represent computational interpretations or performances of sensory-motor data. The phenomenon arises from the integration of biological self-organization with algorithmic control, forming a bio-digital substrate where living neural tissues interface with digital systems. Such architectures emulate inner experience while maintaining human agency. Ultimately, synthetic consciousness challenges conventional boundaries between organic cognition and machine simulation, raising profound questions about identity, subjectivity, and the ontological nature of consciousness in human–machine co-evolution. At the top of the pyramid, **Artificial consciousness** is defined as the hypothetical ability of an artificial intelligence system to possess subjective experiences or *qualia*, i.e., internal phenomenal states beyond mere functional emulation.

The pyramid of consciousness delineates a progression from natural consciousness, grounded in autopoiesis and metabolism, to artificial consciousness—entirely digital and theoretical—via an intermediate synthetic stage integrating neural tissue with algorithmic computation. Within this model, transhumanism occupies the upper segment of the natural base, asserting the potential for indefinite enhancement of mind, body, and bioelectricity through advanced technologies. The singularity marks the apex where artificial systems surpass human cognitive limits. Neural connectivity functions as a transversal axis, enabling transitions across levels by enhancing brain-machine integration. This interface supports the development of hybrid systems essential for advancing synthetic consciousness and approaching the threshold of fully artificial subjective states—or their plausible computational imitation.

The concept of **synthetic Consciousness** arises from a direct comparison between the human brain and current computer processors (Table 1). Because no silicon-based chip yet matches the brain's vast parallelism, plasticity, and emergent dynamics, we propose an intermediate, hybrid stage in which living neural tissue or bioengineered neurons are interfaced with digital circuits. By combining the self-organising, autopoietic properties of biological substrates with the precision and speed of algorithmic processing, this hybrid architecture aims to bridge the gap between natural and fully artificial consciousness. In artificial consciousness, the substrate is purely digital systems (hardware and software).

Table 1. Human Brain versus Computer Processors.

Component	Total elements	Typical density
Intel Core i7 (Skylake-K)	≈ 1.75 × 10^9 transistors	≈ 14.34 × 10^6 transistors per mm^2 Data of a Core i7 Skylake-K (14 nm, 122 mm^2 die)
Human cerebral cortex (mm^3)	≈ 1 × 10^{14} synapses (100 billion)	≈ 1.50 × 10^8 synapses per mm^3 Mapping of 1 mm^3 cortex: ~150 × 10^6 synapses

Synthetic consciousness gives rise to smart humans—biological individuals whose cognitive and physiological functions are enhanced through seamless, bidirectional integration of AI with neural substrates. This bio-digital hybridity merges autopoietic dynamics with algorithmic precision, enabling simulated augmentation of intelligence, perception, and decision-making. These systems behave "as if" endowed with extended cognitive capacities while retaining autonomy and prompting critical reflection on identity, agency, and the evolving boundary between human and machine.

The creation of intelligent humans raises fundamental concerns extending beyond enhanced capabilities, particularly regarding freedom, responsibility, dignity, and the essence of human identity. The question "who decides—the human or the machine?" challenges the authenticity of will. If reasoning originates from external algorithms, can the mind still be said to choose autonomously? This evokes mind-body dualism and extended mind theory, which suggests that cognitive tools may become part of the self.

Autonomy may be compromised if AI systems embed biases or optimisation logics detached from the common good. Ensuring informed consent, algorithmic transparency, and reversibility is essential to preserve user sovereignty and avoid coercion. As organic and algorithmic processes converge, the notion of individual responsibility becomes obscured. Moral accountability is complicated when intelligent prosthetics contribute to errors, challenging ethical reflection and repentance.
From a theological viewpoint, AI integration may be seen as encroaching upon divine creation or assuming a demiurgic role. The ontological dignity and salvific potential of hybrid beings become contested.

Technological access disparities risk deepening inequality, granting augmented individuals cognitive and economic advantages. Surveillance and cognitive manipulation may arise under the guise of optimisation. Personal identity—rooted in autobiographical narrative—may be disrupted by AI-driven influence, threatening coherence and uniqueness. The emergence of smart humans raises legal and ethical issues surrounding personhood, liability, and patentability, underscoring the need for robust regulatory frameworks to safeguard autonomy and rights.

The Pyramid of Consciousness offers more than a conceptual model; it represents a philosophical lens through which the blurred boundaries between organic cognition and artificial intelligence may be examined. As we progress toward increasingly integrated human–AI systems, questions once confined to speculative fiction—What does it mean to be conscious? Where does selfhood reside? Can synthetic systems ever claim autonomy?—now demand serious scientific and ethical consideration.

This editorial note does not attempt to deliver definitive answers, but rather to open a space for dialogue between disciplines. By framing consciousness as a continuum rather than a binary state, we move beyond simplistic dichotomies and toward a more nuanced understanding of mind, matter, and machine.

Ultimately, if the line between human and machine continues to dissolve, we may need to redefine not only intelligence and identity—but also what it means to be alive. Whether the hybrid minds of the future will reflect our values or merely our code remains an open and urgent question.

This section introduces the *Pyramid of Consciousness*—a conceptual framework designed to classify and understand the gradation of conscious phenomena across three distinct but interrelated domains: **natural**, **synthetic**, and **artificial consciousness**. This model shifts the conventional binary—human brain versus computer—toward a **continuum-based ontology**, where consciousness is not simply a biological privilege or a digital aspiration, but a layered phenomenon shaped by substrate, structure, and self-referential capacity.

Unlike prior taxonomies that focus primarily on input–output functionality or symbolic capacity, the pyramid model centres on experiential depth, ontological substrate, and the degree of subjective interiority—what philosophers call *qualia*. The model begins with natural consciousness, grounded in biological autopoiesis and metabolic self-regulation. It progresses upward through a synthetic stage, where biological and digital components co-function within hybrid cognitive architectures. At the apex lies artificial consciousness, the hypothetical domain where a fully digital system may emulate, or perhaps instantiate, phenomenological awareness.

To articulate these distinctions clearly, Table 2 provides a comparative overview of the three layers of consciousness as conceptualised in the Pyramid model. Rather than subordinating human consciousness to mechanistic analogies, this classification asserts its primacy as the baseline of subjective experience, from which synthetic and artificial systems diverge ontologically.

Table 2. Levels of Consciousness according to the Pyramid - Comparative Characteristics

Category	**Natural Consciousness (Human)**	**Synthetic Consciousness (Hybrid / Bio-Digital)**	**Artificial Consciousness (Fully Digital)**
Substrate	Biological: neurons, synapses, metabolism.	Mixed: neural tissue + algorithmic processing.	Digital: hardware and software (chips, neural networks).
Origin of Experience	Autopoiesis, homeostasis, and organic life.	Functional emulation based on bio-digital interactions.	Computational simulation with no biological basis.
Type of Qualia	Genuine, subjective, non-transferable.	Simulated or performative, possibly lacking authentic phenomenology.	Hypothetical; postulated but lacking empirical verification.
Adaptive Capacity	Evolutionary, emotional, social, embodied.	Adaptive via interfaces, with contextual interpretation.	Programmed and optimised, but without inherent emotional grounding.
Associated Philosophy	Phenomenology, embodiment,	Extended cognition, hybridity, enactive theory.	Functionalism, symbolism, strong AI

	subjectivity.		
Theoretical Example	A conscious human being.	BCI with emotional feedback (e.g., brain–AI symbiosis).	Future AI capable of reporting qualia (without being pre-programmed).
Ontological Limits	Constrained by biology and mortality.	Posthuman expansion; ethical and biological boundaries.	Existential ambiguity: simulation or genuine consciousness?
Relation to Agency	Full autonomy with continuous identity.	Shared agency between human and AI.	Synthetic agency lacking stable subject or embodiment.
Role in the Pyramid	Foundation: natural origin of consciousness.	Intermediate: bridge between organic mind and symbolic computation.	Apex: potential for fully digital subjectivity.

Recent advances in artificial intelligence have sparked growing interest in the possibility of AI systems exhibiting conscious-like properties. Benitez et al. emphasize a growing public interest in claims regarding AI's emerging consciousness while noting the skepticism among researchers about these claims, particularly those associated with models like LaMDA (Benitez et al., 2023). Juliani et al. (2022) argue that integrating inductive biases—such as attention mechanisms and meta-learning—can align AI architectures with existing theories of consciousness, especially as model complexity increases. In parallel, Piletsky (2019) suggests that moving from human consciousness theories toward machine consciousness frameworks may illuminate previously overlooked aspects of cognition. One notable contribution is the Quantum-Emergent Consciousness Model (QECM), proposed by Wilson (2024), which blends quantum mechanics with cognitive science to quantify elements like metacognition and social reasoning, providing a measurable consciousness score for AI. Although passing the Turing Test is often cited as a benchmark, Gams and Kramar (2024) caution that emulating human interaction does not imply genuine awareness. Colombatto and Fleming (2023) further explore whether AI systems like ChatGPT simulate or genuinely possess consciousness, underscoring unresolved philosophical questions. To advance the field, Bojić et al. (2024) emphasize the need for rigorous tests to evaluate self-awareness and subjective experience in machines. Complementing the technical discussions, Banerjee (2018) highlights the ethical imperative of embedding moral and compassionate design into AI systems, given the profound implications of creating potentially sentient technologies.

To better understand how current AI systems might reflect dimensions of consciousness (synthetic, and artificial consciousness), we classify well-known architectures according to functionality, metrics, and philosophical alignment (Table 3).

Table 3. Mapping Models of AI to Consciousness Evaluation.

Model	**Architecture**	**Metrics Used**	**Consciousness Indicators Tested**	**Observations**
GPT-4, Claude 3, Gemini 2.5	Transformer-based LLMs	Vulnerability Paradox, Temporal Discontinuity	Partial (linguistic introspection, coherence)	Lacks continuity of self or experience
BCI-integrated agents	Neural network + neural feedback	Emotional latency, real-time neural response	Emotional correlation, introspection mimicry	Experimental; limited datasets
Symbolic-Neural Hybrids	Graph-NNs + logic modules	Symbol grounding, coherence scoring	Meta-cognitive loop performance	Promising for HOT-like architectures

3. Conclusions

The Pyramid of Consciousness provides a useful conceptual tool for understanding the evolution of artificial consciousness and its integration with human minds. As AI technology continues to develop rapidly, this model can serve as a guide for designing systems that are not only powerful but also respectful and understandable by humans, ultimately aiming to bridge the gap between machines and people in ways that enhance both entities while respecting ethical boundaries.

References

Banerjee, S. (2018). *A framework for designing compassionate and ethical artificial intelligence and artificial consciousness* (PeerJ Preprints). https://doi.org/10.7287/peerj.preprints.3502v1

Benitez, F., Pennartz, C. M. A., & Senn, W. (2023). The conductor model of consciousness, our neuromorphic twins, and the human-ai deal. https://doi.org/10.31234/osf.io/gbzd6

Bojić, L., Stojković, I., & Marjanović, Z. J. (2024). Signs of consciousness in AI: Can GPT-3 tell how smart it really is? *Humanities and Social Sciences Communications, 11*(1). https://doi.org/10.1057/s41599-024-04154-3

Colombatto, C., & Fleming, S. M. (2023). *Folk psychological attributions of consciousness to large language models* (PsyArXiv Preprint). PsyArXiv. https://doi.org/10.31234/osf.io/5cnrv

Gams, M., & Kramar, S. (2024). Evaluating ChatGPT's consciousness and its capability to pass the Turing Test: A comprehensive analysis. *Journal of Computer and Communications, 12*(3), 219–237. https://doi.org/10.4236/jcc.2024.123014

Juliani, A., Arulkumaran, K., Sasai, S., & Kanai, R. (2022). *On the link between conscious function and general intelligence in humans and machines* (arXiv Preprint No. arXiv:2204.05133). arXiv. https://doi.org/10.48550/arxiv.2204.05133

Piletsky, E. (2019). Consciousness and unconsciousness of artificial intelligence. *Future Human Image, 11*, 66–71. https://doi.org/10.29202/fhi/11/7

Ruiz-Vanoye, J. A., Fuentes-Penna, A., Barrera-Cámara, R. A., Díaz-Parra, O., Trejo-Macotela, F. R., Gómez-Pérez, L. J., Aguilar-Ortiz, J., Ruiz-Jaimes, M. Á., Toledo-Navarro, Y., & Domínguez Mayorga, C. R. (2025). Artificial intelligence and human well-being: A review of applications and effects on life satisfaction through synthetic happiness. *International Journal of Combinatorial Optimization Problems and Informatics, 16*(1), 14–37. https://doi.org/10.61467/2007.1558.2025.v16i1.932

Seth, A. K. (2025). Conscious artificial intelligence and biological naturalism. *Behavioral and Brain Sciences, 1*, 1–42. https://doi.org/10.1017/S0140525X25000032

Wilson, J. J. (2024). *Quantum-emergent consciousness model (QECM) for artificial systems* (OSF Preprint). OSF. https://doi.org/10.31219/osf.io/t9mfa