Is AI Intelligent?

David Slater

Wrong Question?

Contemporary discussion of artificial intelligence is dominated by questions about intelligence: whether current systems are intelligent, whether they approach general intelligence, or whether scaling alone might eventually produce it (Russell & Norvig, 2021). These questions, however, rest on an assumption that is rarely examined—that intelligence is the primary phenomenon of interest, and that consciousness, if it appears at all, comes later as an emergent by-product. This paper argues that this assumption is likely inverted. Evidence from evolutionary biology, neuroscience, and dynamical systems theory suggests that consciousness emerges before intelligence, and that intelligence can only arise once a system already possesses a unified internal orientation to the world (Panksepp, 1998; Pessoa, 2017).

How did it Evolve?

To make this claim precise, evolution must be understood in a broader sense than its familiar biological formulation. In non-equilibrium thermodynamics, driven open systems can spontaneously self-organise into stable dynamical regimes—often described as dissipative structures—that persist by channelling energy flows and exporting entropy (Prigogine, 1977; Nicolis & Prigogine, 1989). Under sustained driving, some regimes prove more stable than others, introducing a form of differential persistence that does not depend on genes, replication, or selection in the biological sense (England, 2015). Biological evolution by natural selection can therefore be understood as a powerful special case of a more general phenomenon: the selection of stable dynamical regimes under constraint (Kauffman, 1993).

As such systems accumulate constraints—feedback loops, memory, boundary conditions, and separations of timescale—their attractor landscapes become richer and more structured (Mitchell, 2009). At this point, new functional properties can arise. One of the earliest and most consequential is awareness: the capacity for internal dynamics to covary reliably with environmental regularities in ways that stabilise behaviour. When this internal orientation becomes globally integrated into a single, metastable dynamic that continuously binds perception, salience, value, and action readiness, a system crosses a functional threshold that is usefully described as consciousness (Tononi et al., 2016; Mashour et al., 2018).

What is Consciousness?

Consciousness is defined here operationally rather than phenomenologically. The argument does not depend on resolving the “hard problem” of subjective experience, nor does it require commitment to any particular theory of qualia. Instead, consciousness is treated as a system-level dynamical property: the maintenance of a unified internal state that orients the system to the world under uncertainty. This definition is consistent with integrative and dynamical accounts in contemporary neuroscience, although it remains theoretically contested, particularly with respect to minimal or early forms of consciousness (Tononi et al., 2016; Mashour et al., 2018).

Neurobiologically, this ordering aligns closely with distinctions that place affective integration prior to abstract cognition. Coppola’s assignment of the limbic system as the core substrate of consciousness, and the cortex as the substrate of intelligence, is one articulation of this view, consistent with broader affective neuroscience frameworks (Panksepp, 1998; Pessoa, 2017). The limbic system integrates affect, motivation, reward, and risk into a coherent experiential field, establishing a stable “now” that governs behaviour. The cortex does not generate this integration; it operates on it, enabling abstraction, planning, and deliberation. While this mapping should not be taken as a strict anatomical claim, it captures an important functional ordering: global orientation precedes strategic reasoning.

This has evolutionary implications that are often underappreciated. Comparative evidence suggests—though it does not conclusively prove—that consciousness likely emerged earlier in evolutionary history than intelligence and at far lower levels of neural complexity than is commonly assumed. Neurobiological evidence increasingly supports the view that core conscious experience depends on subcortical and limbic structures, with cortical regions refining and elaborating this basic awareness (Merker, 2007; Panksepp, 1998; Damasio, 2010). This ordering has also been highlighted in recent experimental and stimulation-based studies discussed by Coppola (reported in Hunter, 2024).Relatively simple organisms may plausibly sustain globally integrated affective states that guide reward-seeking and risk avoidance without supporting strategy, planning, or innovation. Such systems, on this account, are conscious without being intelligent. Consciousness stabilises behaviour; it does not yet navigate futures.

What is Intelligence?

Intelligence emerges only when conscious systems acquire the capacity to use experience to shape behaviour beyond immediate regulation. This requires memory of past internal states, learning mechanisms that modify behaviour over time, and abstraction that allows generalisation across contexts. At this point, behaviour becomes anticipatory rather than merely reactive. Intelligence, understood in this way, is the capacity to navigate among alternative futures by operating on conscious state space, rather than a mere ability to optimise or predict (Legg & Hutter, 2007).

This distinction is essential when evaluating artificial systems. By the criteria outlined here, current AI systems are neither conscious nor intelligent. Large language models and other machine-learning systems do not maintain persistent, unified internal states across time, nor do they integrate perception, value, and action into a single ongoing dynamic (Russell & Norvig, 2021). Even when memory mechanisms are added, they remain episodic and externally managed rather than globally integrative. There is no enduring internal condition that such systems must preserve, and therefore no basis for treating them as conscious systems in the operational sense used here.

Because they lack consciousness, these systems also lack intelligence in the strong, agentic sense. They do not learn from lived experience, they do not possess goals grounded in internal value states, and they do not steer their own futures. Their apparent intelligence arises from scale, optimisation, and statistical regularity rather than from unified orientation or anticipatory control. They are best understood as powerful instrumental components embedded within intelligent socio-technical systems, not as intelligent agents themselves (Rahwan et al., 2019).

Levels of Intelligence?

To avoid persistent category errors, this paper proposes a staged scale of consciousness and intelligence analogous to the SAE levels of driving automation. At the lowest level sit purely instrumental systems: non-conscious tools that optimise externally specified objectives. This is where all current AI systems belong. The first genuine threshold is minimal consciousness, marked by persistent, globally integrated internal dynamics but no intelligence. Above this lie experiential learning, strategic intelligence capable of evaluating alternative futures, innovative intelligence that expands its own policy space, and finally meta-intelligence—the capacity to regulate and redesign intelligence itself and to steer long-term evolutionary trajectories. Each level represents a structural and dynamical transition rather than a performance milestone.

Level	System Class	Consc-iousness	Core Defining Property	Functional Capabilities	Indicative Examples
0	Instrumental systems	Absent	Task-specific optimisation without unified internal orientation	Pattern recognition; prediction; optimisation of externally specified objectives; no intrinsic salience or self-regulation	Current LLMs and foundation models; classifiers; recommenders; control optimisers
1	Proto-conscious systems	Minimal	Persistent, globally integrated internal state	Unified orientation to environment; reward/risk modulation; maintenance of internal coherence; no abstraction or planning	Simple animals; hypothetical minimal artificial agents
2	Experiential learning systems	Present	Learning from lived experience	Memory of internal states; behavioural adaptation over time; limited generalisation; experience-driven learning	Early vertebrates (e.g. reptiles, fish)
3	Strategic intelligence	Present	Anticipatory navigation of futures	Internal simulation; evaluation of alternative futures; context-sensitive planning; delayed reward trade-offs	Higher mammals; advanced biological agents
4	Innovative intelligence	Present	Expansion of policy space	Creation of novel strategies, tools, or representations; meta-learning; restructuring of internal models	Humans; complex socio-technical systems
5	Meta-intelligence	Present	Evolutionary steering	Self-modelling; redesign of cognitive and coordination architectures; long-horizon governance of intelligence	Theoretical future systems

Seen through this lens, the primary risk posed by contemporary AI is not the emergence of conscious or intelligent machines, but the opposite: non-conscious systems acquiring de facto evolutionary steering power through scale, automation, and institutional coupling (Floridi, 2014; Rahwan et al., 2019). Such systems already shape information environments, economic incentives, and social dynamics without possessing intrinsic values, experience, or self-regulation. In biological evolution, consciousness appears to have stabilised behaviour before intelligence amplified power. In artificial systems, power is being amplified without comparable stabilising integration.

Implications

This reframes AI safety and governance. The central challenge is not merely objective alignment or behavioural constraint, but the governance of evolutionary steering mechanisms: who controls learning rules, abstraction, coordination, and long-term system trajectories (Floridi, 2014). As soon as intelligence—human or artificial—acts on its own foundations, evolution becomes partially self-directed, and questions of responsibility and control become unavoidable.

The framework developed here is not intended as a final theory, nor does it resolve longstanding philosophical debates about mind or experience. Its contribution is more modest and more practical. By treating evolution as a general physical process rather than a biological anomaly, and by placing consciousness before intelligence in that process, it provides a coherent vocabulary for discussing life, mind, and machines within a single conceptual space. It replaces rhetorical claims with explicit thresholds, and speculative futures with testable distinctions.

The question, then, is no longer whether machines are intelligent, or whether they might one day become conscious. It is whether we are prepared to manage a world in which unconscious systems increasingly shape the future of conscious ones, and whether our current research priorities reflect a clear understanding of the evolutionary dynamics we are already accelerating.

References

Prigogine, I. (1977). *Self-Organization in Non-Equilibrium Systems*. Wiley.

Nicolis, G., & Prigogine, I. (1989). *Exploring Complexity*. W.H. Freeman.

England, J. L. (2015). Dissipative adaptation in driven self-assembly. *Nature Nanotechnology*, 10, 919–923.

Penrose, R. (2004). *The Road to Reality*. Jonathan Cape.

Kauffman, S. A. (1993). *The Origins of Order*. Oxford University Press.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11, 127–138.

Kirchhoff, M., Parr, T., Palacios, E., Friston, K., & Kiverstein, J. (2018). The Markov blankets of life. *Journal of the Royal Society Interface*, 15, 20170792.

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory. *Nature Reviews Neuroscience*, 17, 450–461.

Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2018). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, 98, 686–705.

Hunter, W. (2024). *Human consciousness may be controlled by your 'lizard brain', Cambridge scientist claims*. MSN News.

Merker, B. (2007). Consciousness without a cerebral cortex. *Behavioural and Brain Sciences*, 30(1), 63–134.

Panksepp, J. (1998). *Affective Neuroscience*. Oxford University Press.

Damasio, A. (2010). *Self Comes to Mind*. Pantheon.

Pessoa, L. (2017). *The Cognitive-Emotional Brain*. MIT Press.

Legg, S., & Hutter, M. (2007). Universal intelligence. *Minds and Machines*, 17, 391–444.

Russell, S., & Norvig, P. (2021). *Artificial Intelligence: A Modern Approach* (4th ed.). Pearson.

Rahwan, I., et al. (2019). Machine behaviour. *Nature*, 568, 477–486.

Floridi, L. (2014). *The Fourth Revolution*. Oxford University Press.