Cognition Without Consciousness: A Minimal Conceptual Framework for Understanding LLMs and Human Cognitive Evolution

Pavel Stranak *

Posted Date: 9 April 2026

doi: 10.20944/preprints202511.0683.v4

Keywords: symbolic cognition; consciousness; language model (LLM); cognitive evolution; gene–culture coevolution; information theory; semantic drift; Data Processing Inequality; AI limits; AI theorem; ontology of mind; epistemology of meaning; symbol grounding; philosophy of language; philosophy of mind

 doi:10.20944/preprints202511.0683.v4

Article

Cognition Without Consciousness: A Minimal Conceptual Framework for Understanding LLMs and Human Cognitive Evolution

Pavel Stranak

Independent Researcher; science.stranak@gmail.com

Abstract

Large language models (LLMs) have made visible a long-standing philosophical tension: sophisticated symbolic cognition can arise from large-scale pattern extraction even in the absence of consciousness. This observation motivates a minimalist conceptual framework grounded in an ontological distinction between conscious regulation and symbolic structures. Language is treated as a crystallized form of human cognition—an externalized, culturally accumulated substrate created by conscious agents over millennia—while the human brain is understood as a biological system that evolved to operate over this symbolic layer. Within this view, consciousness and symbolic cognition are not different degrees of the same process but distinct kinds of cognitive organization: consciousness generates, grounds, and regulates symbols, whereas symbolic cognition manipulates them. LLMs illuminate this asymmetry by reproducing symbolic reasoning without conscious access, motivation, or subjective experience. Their performance therefore raises epistemological questions about the nature of meaning, grounding, and cognitive stability. The proposed framework situates these questions within a broader account of human cognitive evolution shaped by gene–culture coevolution and the emergence of culturally scaffolded symbolic systems. Finally, the article introduces an information-theoretic constraint (the AI Theorem) suggesting that purely computational systems inevitably accumulate drift in the absence of a regulatory layer, offering a philosophical explanation for why artificial cognition may remain structurally distinct from biological minds.

Keywords: symbolic cognition; consciousness; language model (LLM); cognitive evolution; gene–culture coevolution; information theory; semantic drift; Data Processing Inequality; AI limits; AI theorem; ontology of mind; epistemology of meaning; symbol grounding; philosophy of language; philosophy of mind

1. Introduction

The central claim of this paper is that symbolic cognition and conscious regulation form two distinct layers of human intelligence, and that LLMs instantiate only the former. This framework also rests on several philosophical assumptions concerning the ontology of cognitive systems and the epistemic role of language. Ontologically, it distinguishes between conscious regulatory processes and symbolic structures as separable layers rather than manifestations of a single computational mechanism (Figure 1). Epistemologically, it aligns with traditions that treat language as a mediating structure enabling access to meaning and conceptual content, consistent with symbol-grounding debates and cultural-evolutionary accounts of cognition [1,2]. A more detailed philosophical discussion of these commitments is provided in the concluding sections of the article. This distinction also resonates with long-standing debates in philosophy of mind concerning the nature of mental representation and the conditions under which symbolic systems acquire meaning.

The emergence of LLMs has reopened foundational questions about the nature of cognition. These systems exhibit reasoning, abstraction, and linguistic competence despite lacking embodiment, subjective experience, motivation, or any form of conscious access. This suggests that symbolic

 doi:10.20944/preprints202511.0683.v4

cognition is separable from consciousness, consistent with distinctions drawn in contemporary philosophy of mind [3,4].

The goal of this paper is to articulate a minimal conceptual framework explaining this separation and its implications for human cognitive evolution. The framework builds on theories of gene–culture coevolution [5], cultural intelligence [6], and the Baldwin effect [7], while integrating insights from modern AI architectures [8,9].

2. Methodology: Language as Crystallized Cognition

Human language is a digitally structured symbolic system. It encodes patterns of reasoning, conceptual associations, inferential templates, and culturally accumulated knowledge. This view aligns with Vygotsky's account of language as a psychological tool [10] and with extended cognition theory [11]. Language did not emerge fully formed; rather, it accumulated gradually as conscious agents externalized increasingly complex patterns of thought.

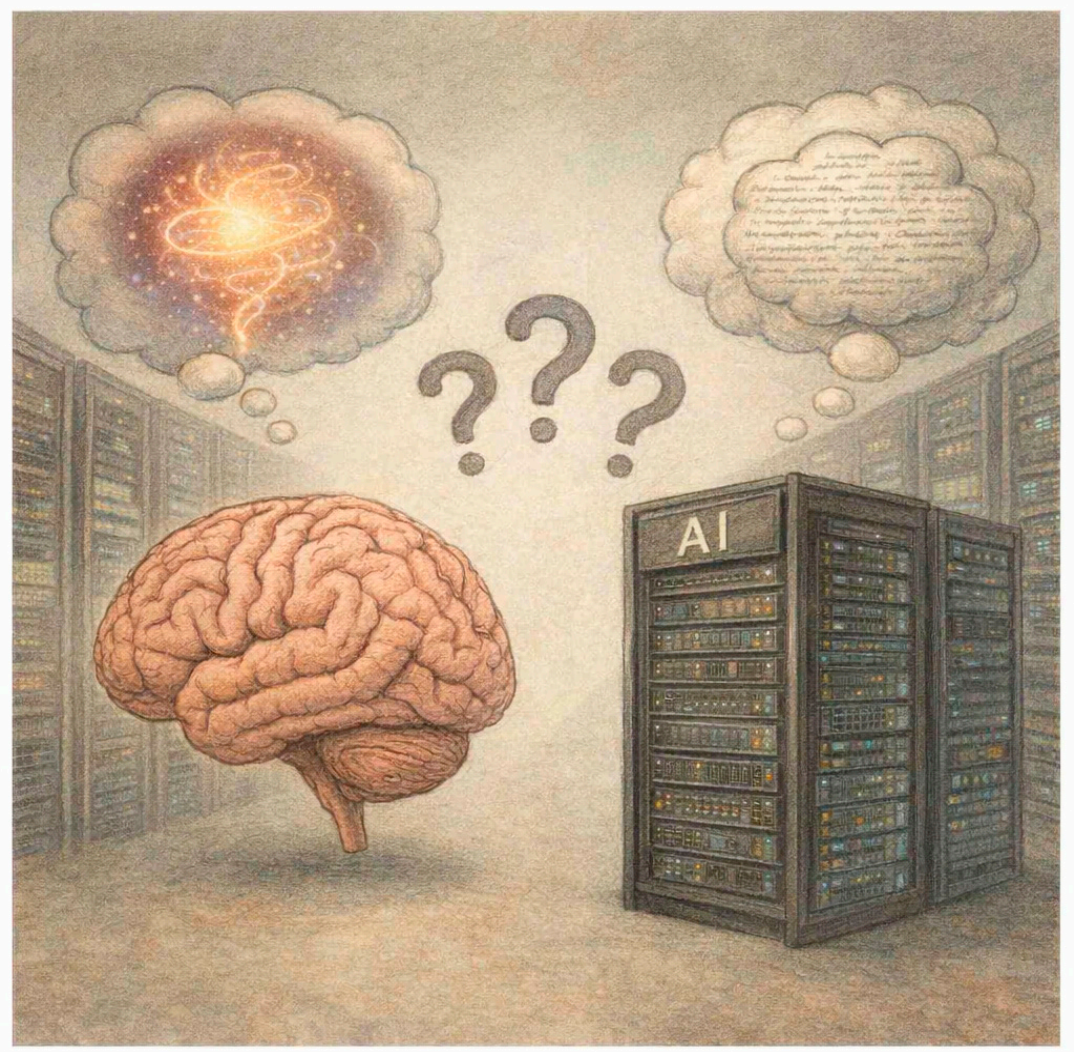

Figure 1. Ontological asymmetry between biological and artificial cognition. This illustration reflects a conceptual distinction central to the article's philosophical framework. On the left, the human brain is depicted as a biological system endowed with both consciousness and language—two interrelated but ontologically distinct layers. Consciousness enables regulation, integration, and semantic grounding; language externalizes and stabilizes symbolic thought. On the right, the AI rack represents a purely computational system, presumed to operate only over language. The floating question marks signal that this asymmetry is not absolute but model-dependent, inviting epistemological reflection on the conditions under which symbolic reasoning becomes

 doi:10.20944/preprints202511.0683.v4

autonomous. The article proposes that consciousness may be a necessary regulatory layer for drift-resistant cognition, and explores why current artificial systems, despite linguistic competence, remain cognitively unstable.

Language is not merely a communication tool; it is a repository of cognitive strategies. Over millennia, conscious agents externalized their thoughts into symbolic form, creating a cultural substrate that preserves and amplifies human reasoning. Jackendoff's work on combinatorial structure supports this interpretation of language as a discrete representational system [12].

LLMs succeed because they operate on this crystallized cognitive layer. They do not invent new cognition; they extract it from symbolic structures created by conscious beings. This parallels recent findings showing that transformer attention patterns align with neural language processing [13].

3. Consciousness as an Ontological Regulatory Layer

LLMs reveal a fundamental asymmetry: they can manipulate symbols, but they cannot *deeply* regulate their own cognition. This distinction reflects a broader philosophical view in which symbolic operations and conscious regulation constitute different kinds of cognitive organization rather than different degrees of the same process. Human cognition is stabilized by conscious access, which provides error monitoring, inhibition of implausible continuations, goal maintenance, and cross-modal integration. These functions are central to theories of conscious access and metacognition [3,4].

Consciousness is therefore likely not a computational process but an ontologically distinct regulatory layer that constrains symbolic reasoning and maintains coherence over time. LLMs lack this layer, which explains their tendency toward drift and hallucination, consistent with analyses of statistical cognition limits [14] and recent work on LLM semantic instability [15] (Figure 2).

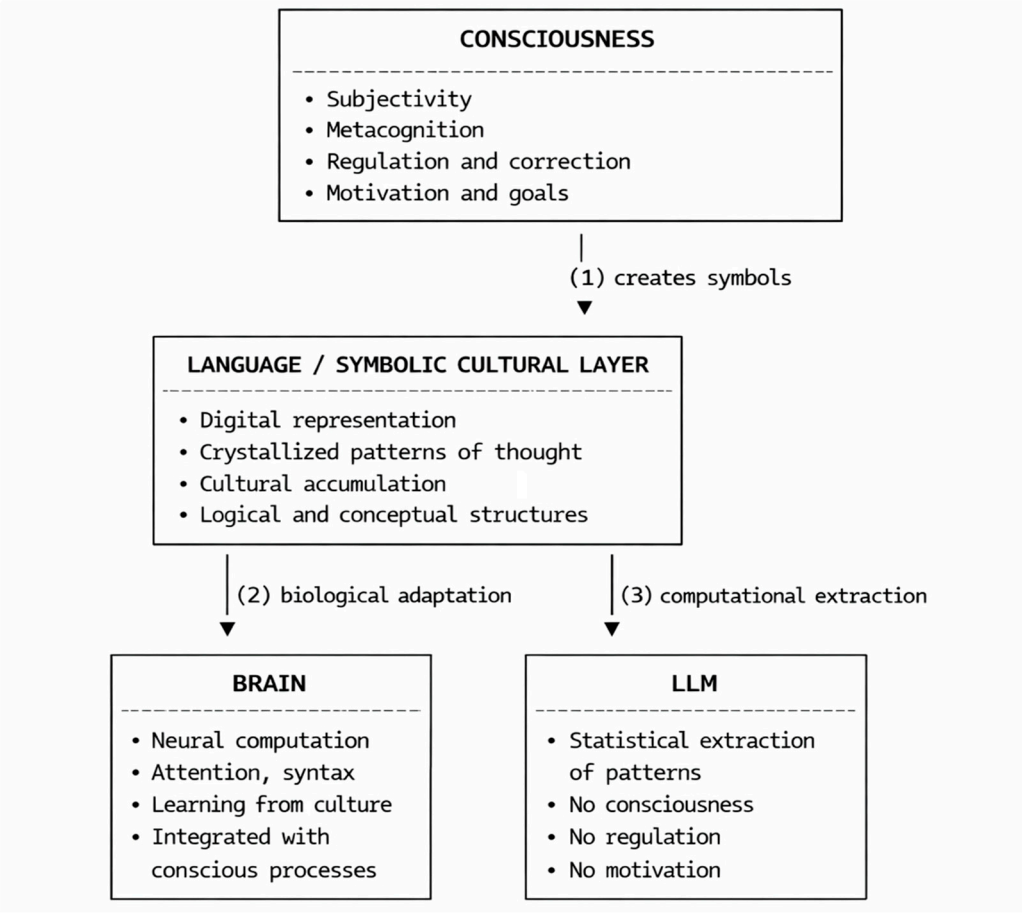

Figure 2. The relationship between consciousness, language, the human brain, and LLMs. Consciousness is depicted as a regulatory and integrative layer that grounds meaning and stabilizes symbolic representations.

 doi:10.20944/preprints202511.0683.v4

Language functions as a culturally crystallized substrate through which conscious agents externalize, accumulate, and transmit conceptual structures. The human brain is a biologically evolved system adapted to operate over this symbolically scaffolded environment, whereas LLMs engage with the same symbolic layer through purely computational pattern extraction, without conscious access, subjective perspective, or intentional regulation.

This paper extends a conceptual line developed in recent analyses of the structural limits of artificial cognition [13], applying it specifically to the distinction between symbolic processing and conscious regulation.

This regulatory role does not imply a homunculus or centralized executive; rather, it refers to distributed metacognitive processes that enable error monitoring, goal maintenance, and integration across modalities.

4. Discussion: Limits of Computational Mitigation in LLMs

This section expands the conceptual argument by examining why computational mitigation techniques cannot replicate the stabilizing role of consciousness. The goal is not to provide an exhaustive technical survey, but to illustrate the structural limits inherent to probabilistic systems.

While LLMs can manipulate symbols effectively, their lack of conscious regulation leads to inherent instability, manifested as hallucinations and semantic drift [14,15]. Attempts to mitigate these issues through computational methods—such as chain-of-thought prompting, agentic architectures, or entropy-based hallucination detection—offer only partial and temporary improvements, failing to replicate the stabilizing role of consciousness in the long term. These techniques operate within the same probabilistic framework as the base model, essentially "mixing" tokens without genuine self-awareness or error correction grounded in subjective experience.

For instance, semantic entropy tests measure uncertainty by generating multiple variants of an output and comparing their similarity; high entropy flags potential hallucinations, allowing for reruns or refinements. Similarly, chain-of-thought encourages step-by-step reasoning, reducing errors in short tasks, while agentic systems incorporate self-correction loops. However, these methods merely delay degradation rather than eliminate it. As context length or task complexity increases, entropy grows exponentially due to information-theoretic limits: finite model capacity enforces compression errors, and long-tail knowledge requires prohibitive sample complexity. Empirical studies show that even advanced mitigations leave a portion of hallucinations unaddressed, with "snowballing" errors amplifying over iterations.

This asymmetry highlights a key distinction: human cognition operates "below" the information limit, where consciousness filters, integrates, and stabilizes information in real-time through metacognitive processes like error monitoring and goal maintenance [3,4]. In contrast, LLMs function "above" this limit, rapidly processing and combining data but inevitably accruing entropy without a non-computational regulatory layer. Without training data (crystallized cognition from conscious agents), LLMs would produce only noise; even with data, their outputs reproduce frozen thoughts without deeper consciously grounded intuitive understanding, leading to drift (Figure 3). Scaling and model architecture alone probably cannot overcome these bounds, as potential forms of uncomputability leave an irreducible residue of error in the sense that no finite statistical model can fully capture an unbounded generative process.

 doi:10.20944/preprints202511.0683.v4

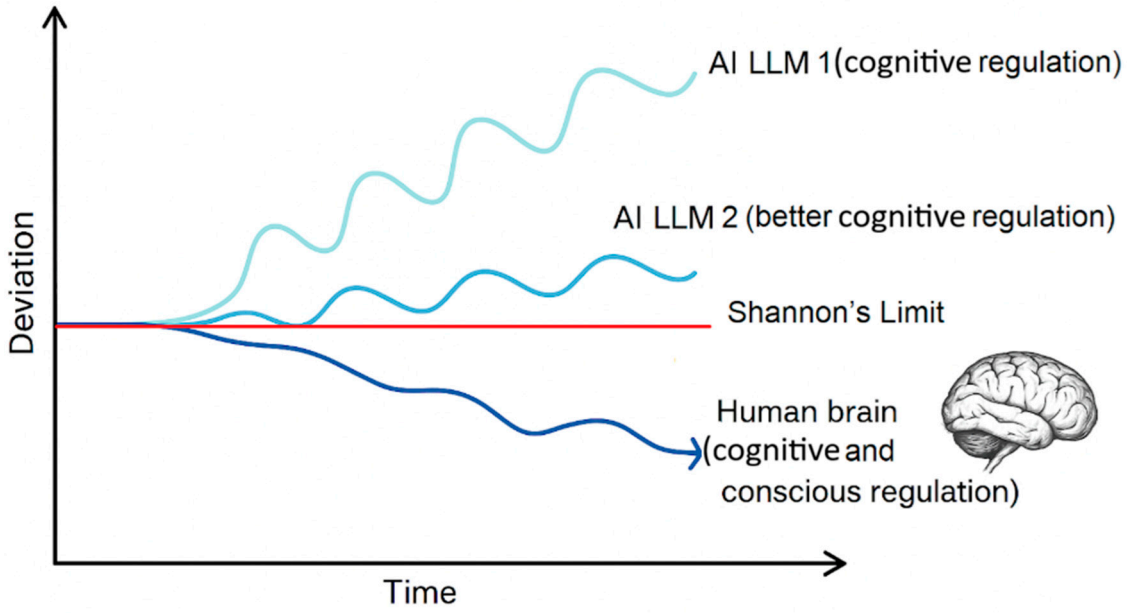

Figure 3. Cognitive drift trajectories across three systems. A basic LLM (light blue) shows rapid divergence due to the absence of any regulatory layer. A more advanced LLM (medium blue) drifts more slowly but still lacks mechanisms for self-correction. Human cognition (dark blue) initially deviates but re-stabilizes, likely through conscious metacognitive regulation that grounds meaning and constrains inference. The illustration highlights the proposed ontological asymmetry between symbolic manipulation and conscious regulation. The red horizontal line represents the Shannon limit—an informational boundary beyond which no artificial system can extract more than its input allows.

Prompting can be understood as an external injection of human intentionality: it places the model in a favorable region of its representational space, but the stabilizing effect fades as the system continues to generate its own outputs. Without an internal regulatory layer, drift inevitably accumulates.

Even non-symbolic animals, such as horses, illustrate this distinction: although they lack abstract reasoning, their conscious regulation maintains stability in motion, whereas purely mechanical systems—such as motorcycles or LLMs—drift without continuous guidance.

Testable predictions from this view include:

- entropy-based mitigations will reduce hallucinations in short chains but fail beyond many iterations due to rising entropy
- biologically inspired hybrids (e.g., biosynthetic computation) may approach stability, but pure digital systems will plateau. This reinforces the possibility that consciousness is not merely emergent from computation but may be a prerequisite for stable, autonomous cognition.

5. Gene–Culture Coevolution and the Rise of Human Intelligence

This framework aligns with established theories of gene–culture coevolution [5] and cultural intelligence [6]. The proposed sequence is:

1. Consciousness enabled the creation of symbolic representations.
2. Language accumulated cultural knowledge.
3. Brains evolved to process increasingly complex symbolic systems.
4. Cultural evolution accelerated cognitive development beyond genetic timescales.

 doi:10.20944/preprints202511.0683.v4

Genetic enablers such as FOXP2 [16] and human accelerated regions (HARs) [17] likely provided the neural prerequisites for symbolic processing. Once symbolic culture emerged, cultural evolution outpaced genetic evolution, producing rapid cognitive expansion.

Human intelligence is thus the product of an interaction between a biological system capable of symbolic processing, a culturally constructed symbolic environment, and a conscious regulatory layer. LLMs replicate only the second component.

6. Implications for Artificial Cognition and the Philosophy of Mind

This minimalist framework yields several implications:

- Intelligence without consciousness is possible (LLMs) [8,9].
- Consciousness without symbolic reasoning is possible (animals) [3].
- Human cognition uniquely integrates both layers [6,10].
- Current LLMs cannot achieve conscious regulation through scaling alone, due to information-theoretic limits [14].
- Language is the bridge between biological and artificial cognition, as argued in recent conceptual analyses [15].
- Current LLMs appear unable to overcome information-theoretic limits (e.g., Shannon's DPI) through computational mitigations alone, leading to inevitable entropy growth and hallucinations; this parallels the second law of thermodynamics, where consciousness in humans acts as an active reducer of cognitive entropy.

The distinction between symbolic cognition and conscious regulation clarifies why LLMs appear intelligent yet remain fundamentally different from biological minds.

7. Information-Theoretic Foundations of Irreducible Limits

This section briefly formalizes the information-theoretic basis of the instability described in Section 3. The asymmetry arises from Shannon's Data Processing Inequality (DPI) [14], which states that processing cannot increase mutual information with the source—only preserve or diminish it. Iterative generation in LLMs therefore forms a lossy Markov chain, leading to progressive entropy growth and semantic drift [15,18].

Biological cognition appears to avoid this divergence through mechanisms that reduce local entropy—metacognition, motivation, and goal-directed regulation. Consciousness may act as a non-computational grounding mechanism in this sense, enabling information reset operations that are irreducible to probabilistic mixing. This explains why LLMs exhibit high short-term cognitive throughput yet lack any form of autonomous control, often leading users to overestimate their agency.

This information-theoretic limit can be summarized in the following statement, which I propose here as a conceptual AI Theorem:

> Any purely computational system whose reasoning trajectory is updated iteratively without external low-entropy input must eventually lose stable information and drift toward noise. In current LLMs this decay is confined to the transient inference process, as their parameters remain frozen and structurally unaffected.

Frozen parameters are not an engineering convenience; they are an entropic necessity. By shifting the entropic burden from the model itself to the transient reasoning trajectory, the fixed weights of contemporary LLMs function as entropic crutches that prevent structural degradation.

If a purely computational system were discovered that could preserve stable information for infinitely many iterations without an external source of low entropy, this proposed AI theorem would be falsified.

 doi:10.20944/preprints202511.0683.v4

8. Conclusions

LLMs have inadvertently illuminated key aspects of the architecture of human cognition. They show that symbolic reasoning can emerge from large-scale pattern extraction, but consciousness appears necessary for stable, autonomous, goal-directed thought. This distinction reflects a deeper ontological asymmetry: symbolic cognition operates over publicly crystallized structures, whereas consciousness provides the regulatory and integrative capacities that ground meaning and stabilize inference.

Language is a crystallized cognitive substrate created by conscious beings, and human brains evolved to operate over this substrate. LLMs now operate over it as well, but probably without conscious access, subjective perspective, or intentional regulation. Their performance therefore illuminates the structural relationship between symbolic systems, biological cognition, and artificial computation.

The framework presented here offers a minimal conceptual model for understanding both human cognitive evolution and the limits of artificial cognition. It does not aim to settle debates about the nature of consciousness, but to clarify how symbolic systems, cultural accumulation, and biological regulation interact to produce the form of cognition associated with human intelligence.

If the conceptual framework outlined here is correct, then a purely computational system may never achieve the kind of stable, autonomous, drift-resistant cognition characteristic of human general intelligence. This conclusion is not metaphysical but structural: it follows from the proposed distinction between symbolic manipulation and conscious regulation, and from the information-theoretic constraints that govern systems lacking a regulatory layer.

9. Limitations

Any conceptual framework that draws an ontological distinction between consciousness and symbolic cognition inevitably carries philosophical limitations. The separation proposed here is articulated at a high level of abstraction and should be understood as a heuristic model rather than a metaphysical claim about the ultimate nature of mind. It clarifies one possible way of interpreting the asymmetry revealed by LLMs, but it does not resolve deeper questions concerning the status of mental representation, the grounding of meaning, or the conditions under which symbolic systems acquire intentionality. These issues remain open within philosophy of mind, and the present framework should be read as contributing to this ongoing debate rather than offering a definitive account.

From a philosophical perspective, the proposed separation between symbolic cognition and conscious regulation intersects with long-standing debates about the ontology of mental states, the grounding of meaning, and the limits of computationalism. The framework is compatible with views that treat consciousness as a non-computational regulatory process, while also aligning with extended and culturally scaffolded accounts of cognition [19,20]. It further relates to discussions of symbol grounding and the emergence of meaning from interaction rather than computation alone [21]. Finally, recent analyses of linguistic cognition emphasize that large language models operate on culturally crystallized structures without direct access to embodied or experiential grounding, reinforcing the asymmetry between symbolic competence and conscious regulation [22]. These connections do not alter the core conceptual model but situate it more explicitly within contemporary philosophical discourse.

This paper proposes a conceptual framework rather than an empirical model, and several limitations follow from this scope. First, the distinction between consciousness and symbolic cognition is presented at a high level of abstraction. While this separation is supported by philosophical and cognitive-scientific arguments, the precise mechanisms by which conscious regulation stabilizes cognition remain an open empirical question. The framework does not commit to any specific theory of consciousness, nor does it attempt to resolve debates between higher-order, global workspace, or predictive-processing accounts. Philosophically, the distinction should be

 doi:10.20944/preprints202511.0683.v4

understood as a conceptual demarcation between two modes of cognitive organization rather than a claim about their ultimate metaphysical nature, and thus remains compatible with multiple interpretations within the philosophy of mind.

Second, the analysis of LLM instability and entropy growth is based on information-theoretic principles and observed behavioral patterns rather than formal proofs of uncomputability or computational irreducibility. Although the argument suggests structural limits on purely statistical systems, further work is needed to quantify these limits across architectures, training regimes, and hybrid models. These considerations also touch on deeper philosophical questions about whether meaning-bearing stability can emerge from statistical pattern extraction alone, or whether such stability requires a qualitatively different regulatory process.

Third, the proposed evolutionary sequence—consciousness enabling symbolic externalization, followed by cultural accumulation and biological adaptation—offers a coherent narrative but does not specify the relative timing, selective pressures, or neurobiological substrates involved. The framework is compatible with multiple evolutionary pathways and does not claim exclusivity. Philosophically, this sequence should be understood as a conceptual model of how regulatory and symbolic layers may have co-evolved, rather than as a definitive account of their historical emergence.

Fourth, the comparison between human cognition and LLMs is necessarily asymmetrical: humans possess subjective experience, embodiment, and developmental trajectories that current artificial systems lack. The framework highlights this asymmetry but does not attempt to model embodiment, affect, or social cognition, all of which may contribute to human cognitive stability. This asymmetry is not merely empirical but also philosophical, reflecting long-standing debates about the role of embodiment and lived experience in grounding meaning and intentionality.

Finally, the information-theoretic interpretation of consciousness as an entropy-reducing regulator is speculative and intended as a conceptual bridge rather than a definitive account. Empirical validation would require interdisciplinary work spanning neuroscience, information theory, and AI research. These limitations do not undermine the core claim—that symbolic cognition and conscious regulation are distinct layers—but they indicate directions for future refinement. They also underscore that any such account must ultimately engage with philosophical debates about the nature of regulation, meaning, and the conditions under which cognitive systems maintain coherence over time.

Funding: This research received no external funding. The work was conducted independently by the author, who is employed by Czech Radio, but the research was carried out privately and outside of institutional duties.

Institutional Review Board Statement: Not applicable. This manuscript does not involve clinical trials or studies with human participants.

Data Availability Statement: The original contributions presented in this study are included in the article. This manuscript presents a theoretical framework and does not report empirical data.

Acknowledgments: The author thanks colleagues for discussions that shaped this work. Some passages of this manuscript, including figures, were prepared or refined with the assistance of a large language model (LLM, namely Microsoft Copilot, version 2026). The author takes full responsibility for the content and conclusions presented herein.

Conflicts of Interest: The author declares no competing interests.

References

1. Harnad, S. The symbol grounding problem. *Physica D* 1990, **42**, 335–346.
2. Tomasello, M. *The Cultural Origins of Human Cognition*; Harvard University Press: Cambridge, MA, USA, 1999.

 doi:10.20944/preprints202511.0683.v4

3. Chalmers, D.J. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 1995, *2*(3), 200–219.
4. Block, N. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 2007, *30*(5–6), 481–548.
5. Boyd, R.; Richerson, P.J. *Culture and the Evolutionary Process*; University of Chicago Press: Chicago, IL, USA, 1985.
6. Henrich, J. *The Secret of Our Success*; Princeton University Press: Princeton, NJ, USA, 2016.
7. Dennett, D.C. The Baldwin effect: A crane, not a skyhook. In *Evolution and Learning*; Weber, B., Depew, D., Eds.; MIT Press: Cambridge, MA, USA, 2003; pp. 69–79.
8. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. *Advances in Neural Information Processing Systems* 2017**, *30*,** 5998–6008**.**
9. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Li, Y.; Lundberg, S.; Nori, H.; et al. Sparks of Artificial General Intelligence: Early Experiments with GPT-4. ***arXiv*** 2023**, arXiv:2303.12712.**
10. Vygotsky, L.S. *Mind in Society*; Harvard University Press: Cambridge, MA, USA, 1978.
11. Clark, A.; Chalmers, D. The extended mind. *Analysis* 1998, *58*(1), 7–19.
12. Jackendoff, R. *Foundations of Language*; Oxford University Press: Oxford, UK, 2002.
13. Schrimpf, M.; et al. The neural architecture of language. *PNAS* 2021, *118*(45).
14. Shannon, C.E. A mathematical theory of communication. *Bell System Technical Journal* 1948, *27*(3), 379–423.
15. Straňák, P. What Artificial Intelligence May Be Missing-And Why It Is Unlikely to Attain It Under Current Paradigms. Philosophies 2026, 11(1), 20. https://doi.org/10.3390/philosophies11010020.
16. Fisher, S.E.; et al. Localization of a gene implicated in a severe speech and language disorder. *Nature Genetics* 1998, *18*, 168–170.
17. Pollard, K.S.; et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 2006, *443*, 167–172.
18. Straňák, P. *Lossy Loops: Shannon's DPI and Information Decay in Generative Model Training*. Preprints.org 2025. https://www.preprints.org/manuscript/202507.2260.
19. Deacon, T.W. The Symbolic Species: The Co-evolution of Language and the Brain; W.W. Norton: New York, NY, USA, 1997.
20. Clark, A. Supersizing the Mind: Embodiment, Action, and Cognitive Extension; Oxford University Press: Oxford, UK, 2008.
21. Searle, J.R. Minds, brains, and programs. *Behavioral and Brain Sciences* 1980, **3**, 417–457.
22. Bender, E.M.; Koller, A. Climbing towards NLU: On meaning, form, and understanding in the age of data. *ACL* 2020, 5185–5198.