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

Pavel Stranak *

Posted Date: 9 February 2026

doi: 10.20944/preprints202511.0683.v2

Keywords: symbolic cognition; consciousness; language model (LLM); cognitive evolution; gene–culture coevolution; information theory; semantic drift; Data Processing Inequality

 doi:10.20944/preprints202511.0683.v2

Article

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

Pavel Straňák

Independent Researcher, Czech Republic, science.stranak@gmail.com

Abstract

Large language models (LLMs) demonstrate that sophisticated symbolic cognition can emerge from scaled pattern extraction without consciousness. This observation motivates a minimalist conceptual framework: language is a crystallized form of human cognition, created by conscious agents over millennia, and the human brain evolved to operate efficiently over this symbolic substrate. Consciousness and symbolic cognition are therefore distinct: consciousness *creates* symbols, while symbolic cognition *operates* over them. LLMs reveal this asymmetry by reproducing symbolic reasoning without possessing conscious regulation, motivation, or subjective experience. This framework clarifies the relationship between biological and artificial cognition and offers a simple model of how human intelligence emerged through gene–culture coevolution.

Keywords: symbolic cognition; consciousness; language model (LLM); cognitive evolution; gene–culture coevolution; information theory; semantic drift; Data Processing Inequality

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.

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 cognition is separable from consciousness, consistent with distinctions drawn in contemporary philosophy of mind [1,2].

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 [3], cultural intelligence [4], and the Baldwin effect [5], while integrating insights from modern AI architectures [6,7].

2. 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 [8] and with extended cognition theory [9]. Language did not emerge fully formed; rather, it accumulated gradually as conscious agents externalized increasingly complex patterns of thought.

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 [10].

 doi:10.20944/preprints202511.0683.v2

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 [11].

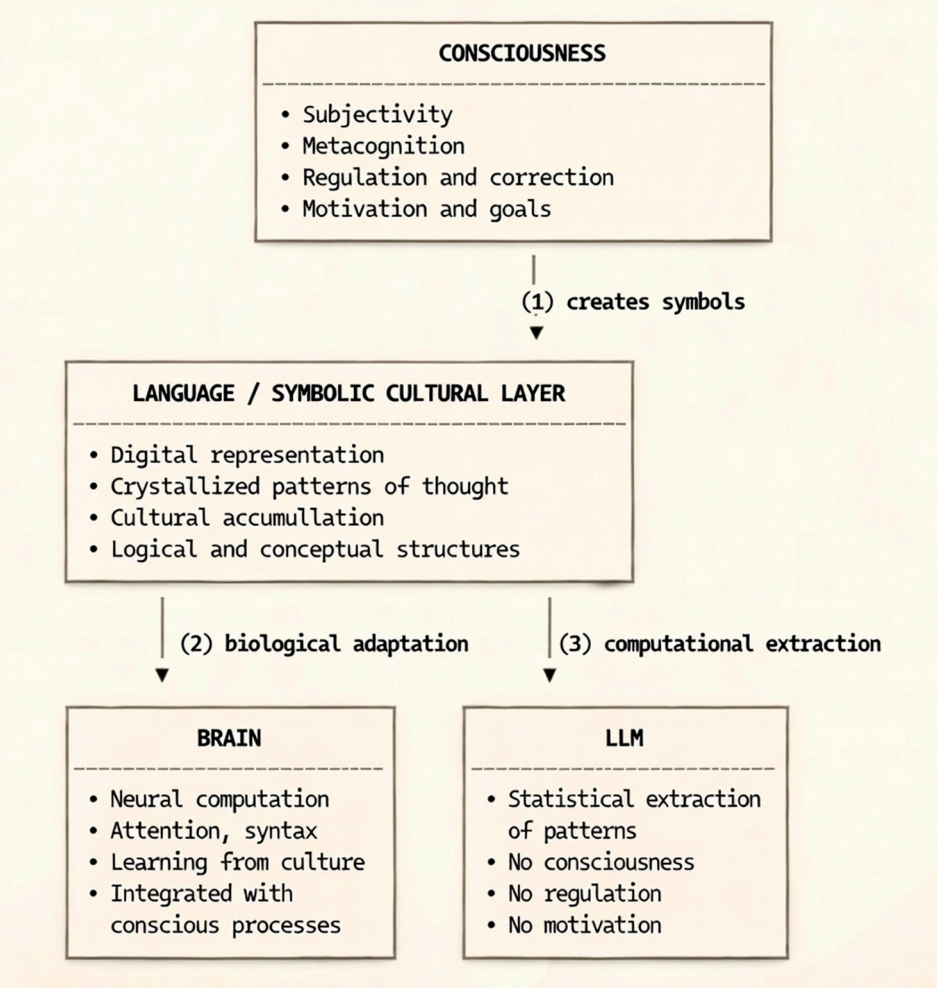

Figure 1. The relationship between consciousness, language, the human brain, and LLMs. Consciousness creates and regulates symbolic representations; language serves as a crystallized cultural substrate; the brain processes these symbols through biologically evolved mechanisms; and LLMs perform purely computational extraction from the same symbolic layer without conscious access.

3. Consciousness as a Regulatory Layer

LLMs reveal a fundamental asymmetry: they can manipulate symbols, but they cannot regulate their own cognition. 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 [1,2].

Consciousness is therefore not a computational process but a 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 [12] and recent work on LLM semantic instability [13].

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.

 doi:10.20944/preprints202511.0683.v2

Limits of Computational Mitigation in LLMs

This subsection 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 [12,13]. 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. 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 [1,2]. 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 understanding, leading to drift, Figure 2. Scaling alone cannot overcome these bounds, as uncomputability ensures an irreducible residue of error in the sense that no finite statistical model can fully capture an unbounded generative process.

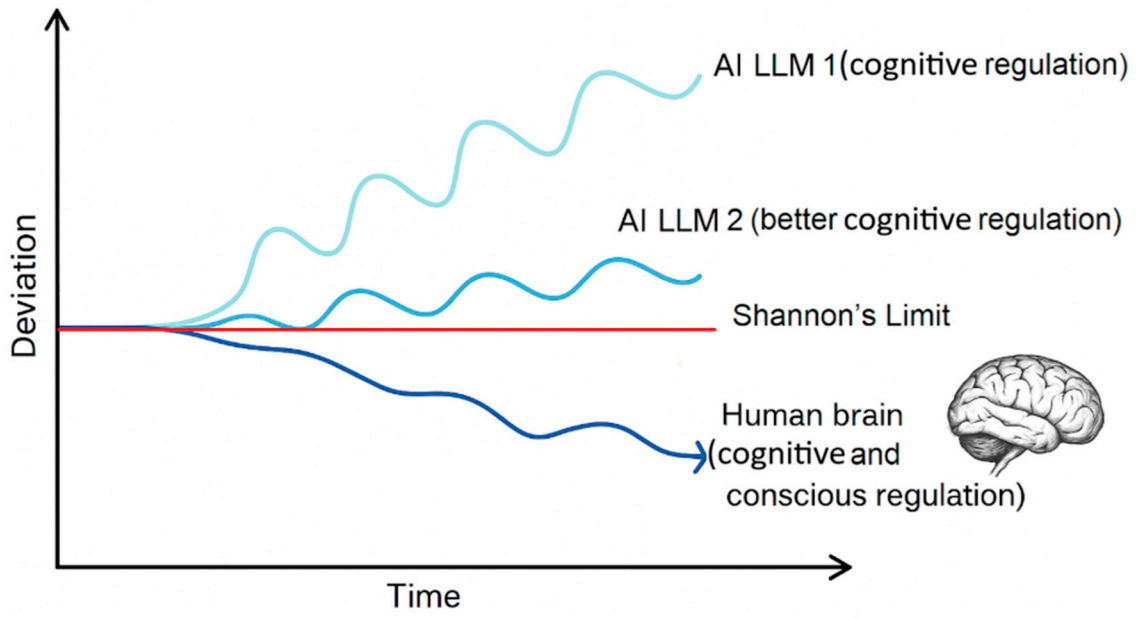

Figure 2. Cognitive drift trajectories across three systems. A basic LLM (light blue) shows rapid divergence due to lack of regulation. A more advanced LLM (medium blue) drifts more slowly. Human cognition (dark blue) initially deviates but re-stabilizes, likely via conscious metacognitive regulation. The illustration highlights the

 doi:10.20944/preprints202511.0683.v2

plausible role of consciousness in maintaining coherence during autonomous reasoning. The red horizontal line represents the Shannon limit—an informational ceiling 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: they lack abstract reasoning, yet their conscious regulation keeps them stable in motion, whereas a purely mechanical system—like a motorcycle or an LLM—rapidly drifts 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 that consciousness is not emergent from computation but a prerequisite for stable, autonomous cognition.

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

This framework aligns with established theories of gene–culture coevolution [3] and cultural intelligence [4]. 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.

Genetic enablers such as FOXP2 [14] and human accelerated regions (HARs) [15] 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.

5. Implications for AI and Philosophy of Mind

This minimalist framework yields several implications:

- Intelligence without consciousness is possible (LLMs) [6,7].
- Consciousness without symbolic reasoning is possible (animals) [1].
- Human cognition uniquely integrates both layers [4,8].
- LLMs cannot achieve conscious regulation through scaling alone, due to information-theoretic limits [12].
- Language is the bridge between biological and artificial cognition, as argued in recent conceptual analyses [13].
- LLMs cannot 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.

 doi:10.20944/preprints202511.0683.v2

6. Information-Theoretic Foundations of Irreducible Limits

Section 6 formalizes the information-theoretic basis of the instability described in Section 3.1.

The asymmetry is rooted in Shannon's Data Processing Inequality (DPI) [12], which states that processing cannot increase mutual information with the source—only preserve or diminish it. Iterative generation in LLMs forms a lossy Markov chain, leading to progressive entropy growth and semantic drift [13,16]. Consciousness acts as a non-computational "information reset," enabling local entropy reduction through metacognition and motivation—mechanisms irreducible to probabilistic mixing. This explains the high short-term cognitive throughput of LLMs juxtaposed with zero conscious control, often leading to overestimation of their autonomy.

7. Conclusions

LLMs have unintentionally clarified the architecture of human cognition. They show that symbolic reasoning can emerge from pattern extraction, but consciousness is required for stable, autonomous, goal-directed thought.

Language is a crystallized cognitive substrate created by conscious beings. Brains evolved to operate over this substrate. LLMs now operate over it too — but without consciousness.

This framework offers a simple conceptual model for understanding both human cognitive evolution and the limits of artificial cognition.

This framework does not aim to settle debates about consciousness, but to clarify the structural relationship between symbolic systems, biological cognition, and artificial computation.

If the conceptual framework outlined here is correct, then a purely computational system may never achieve the kind of stable, autonomous, drift-resistant cognition associated with human general intelligence.

8. Limitations

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.

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.

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.

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.

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.

 doi:10.20944/preprints202511.0683.v2

Funding: This research received no external funding. It was undertaken solely due to the author's personal interest and initiative.

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

Data Availability Statement: Not applicable. This manuscript does not report on 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). The author takes full responsibility for the content and conclusions presented herein.

Conflicts of Interest: The author declares no competing interests.

References

1. Chalmers, D.J. Facing up to the problem of consciousness. *Journal of Consciousness Studies* 1995, *2*(3), 200–219.
2. Block, N. Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behavioral and Brain Sciences* 2007, *30*(5–6), 481–548.
3. Boyd, R.; Richerson, P.J. *Culture and the Evolutionary Process*; University of Chicago Press: Chicago, IL, USA, 1985.
4. Henrich, J. *The Secret of Our Success*; Princeton University Press: Princeton, NJ, USA, 2016.
5. 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.
6. Vaswani, A.; et al. Attention is all you need. In *Advances in Neural Information Processing Systems*; Curran Associates: Red Hook, NY, USA, 2017; Volume 30.
7. Bubeck, S.; et al. Sparks of artificial general intelligence: Early experiments with GPT-4. *arXiv* 2023, arXiv:2303.12712.
8. Vygotsky, L.S. *Mind in Society*; Harvard University Press: Cambridge, MA, USA, 1978.
9. Clark, A.; Chalmers, D. The extended mind. *Analysis* 1998, *58*(1), 7–19.
10. Jackendoff, R. *Foundations of Language*; Oxford University Press: Oxford, UK, 2002.
11. Schrimpf, M.; et al. The neural architecture of language. *PNAS* 2021, *118*(45).
12. Shannon, C.E. A mathematical theory of communication. *Bell System Technical Journal* 1948, *27*(3), 379–423.
13. Straňák, P. *Cognition Without Consciousness: AI Transformers and the Revival of Human Thought*. Preprints.org 2025, doi:10.20944/preprints202511.0683.v1.
14. Fisher, S.E.; et al. Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics* 1998, *18*, 168–170.
15. Pollard, K.S.; et al. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 2006, *443*, 167–172.
16. Straňák, P. *Lossy Loops: Shannon's DPI and Information Decay in Generative Model Training*. Preprints.org 2025, https://www.preprints.org/manuscript/202507.2260.