Invariant-First AI: Compression-Based Architectures for Coherent and Energy-Efficient Artificial Intelligence

Author: Nickolas Patrick Joseph Schoff

Abstract

Contemporary artificial intelligence systems achieve impressive performance at the cost of extreme computational, energetic, and data inefficiency. Large-scale models rely on brute-force accumulation of instances rather than principled abstraction, resulting in instability, hallucination, and escalating resource demands. Drawing on the Unified Consciousness Substrate Theory (UCST) and internal project research (Memory Bank), this paper proposes an alternative design paradigm: invariant-first artificial intelligence. Inspired by single-timeline compression models of reality, we argue that coherent intelligence emerges from the preservation of compressed structural constraints rather than exhaustive memory of events. We formalize a three-layer AI architecture that separates invariant constraint learning from contextual modeling and linguistic expression, demonstrate how this approach reduces computational cost and instability, and map the framework onto information theory and thermodynamics. Implications for AI safety, alignment, and long-term sustainability are discussed.

Keywords: artificial intelligence,

information compression, invariants, coherence, free energy, UCST

1. Introduction

Modern artificial intelligence has entered an era of diminishing returns. Performance improvements increasingly require exponential increases in model size, training data, and energy consumption. These costs are not incidental; they arise from a foundational design assumption that intelligence emerges from large-scale statistical aggregation of surface-level data. While effective, this approach lacks structural memory, leading to incoherence, brittleness, and high operational expense. Parallel research within the Unified Consciousness Substrate Theory (UCST) framework suggests an alternative model of intelligence rooted in compression, constraint, and coherence. UCST posits that reality itself advances through the selection of a single coherent trajectory while compressing unrealized possibilities into invariant structural memory. This paper extends that insight to artificial intelligence, proposing that AI systems should learn and preserve invariants before learning instances. We argue that such systems can achieve greater coherence, predictive power, and efficiency while reducing resource consumption.

2. Background: Compression, Coherence, and UCST

UCST conceptualizes consciousness and intelligence as emergent properties of recursive information integration under constraint. Central to this framework are three principles derived from internal project research (Memory Bank files):

- Constraint precedes form.
- Coherence is achieved through recursive integration across scales.
- Compression preserves structural truth better than accumulation.

In this view, memory is not primarily a record of events but a repository of compressed invariants—structural biases that shape future trajectories. These principles align with established theories in information science, including minimum description length (Rissanen, 1978), the free-energy principle (Friston, 2010), and thermodynamic limits on computation (Landauer, 1961).

3. Limitations of Instance-Based AI Architectures

3.1 Instance Accumulation and Redundancy

Transformer-based models learn correlations across vast corpora of data, repeatedly encoding similar patterns in different forms. Let D represent the training dataset, consisting of instances d_i:

$$D = \{d_1, d_2, ..., d_n\}$$

Where many d_i encode redundant structural information. Compression occurs implicitly and inefficiently across billions of parameters, resulting in large model size and high energy cost.

3.2 Context Reconstruction Cost

At inference time, such models reconstruct coherence dynamically through token-level attention mechanisms. This repeated reconstruction imposes ongoing computational expense and increases the risk of incoherent outputs when context windows are exceeded.

3.3 Structural Amnesia

Because invariants are not explicitly stored, models lack stable long-term constraints. This contributes to hallucination, drift, and the need for frequent retraining or external alignment mechanisms.

4. Invariant-First AI Architecture

We propose a three-layer architecture inspired by UCST's compression hierarchy.

4.1 Layer 1: Structural Invariants (Compressed Core)

This layer stores minimal, slowly changing constraints derived from compression rather than labels. Examples include causal asymmetries, conservation principles, social dynamics invariants, and stability conditions.

Formally, let Ω denote the space of possible models consistent with observed data. Compression yields a set of invariants I:

$$I = \text{min\_code}(\Omega)$$

Where min_code denotes the minimum description length encoding. This layer is not directly queryable and cannot generate outputs; it only biases downstream processes.

4.2 Layer 2: Contextual Models

Contextual models operate under the priors imposed by I, learning domain-specific relationships with significantly reduced data requirements. Given context C and invariants I, predictions P are generated as:

$$P = \text{argmax\_P} \Pr(P \mid C, I)$$

This reduces search space and improves generalization.

4.3 Layer 3: Expression Interface

Language and action are confined to the expression layer. This separation ensures that linguistic fluency does not substitute for structural understanding and prevents the invariant layer from being misinterpreted as propositional knowledge.

5. Thermodynamic and Information-Theoretic Mapping

5.1 Entropy Redistribution

Total entropy must be conserved:

$\Delta S\_total \geq 0$

Invariant-first AI reduces local entropy in contextual reasoning by exporting entropy into the compressed invariant layer:

$\Delta S\_total = \Delta S\_context + \Delta S\_invariant$
This mirrors single-timeline compression models in UCST.

5.2 Free Energy Minimization

Free energy F is defined as:

$F = E - T S$

Invariant priors reduce surprise and uncertainty, allowing the system to minimize F more efficiently. Predictions that violate deep constraints are penalized early, reducing wasted computation.

5.3 Energetic Cost of Information Erasure

Landauer's principle states that erasing one bit of information requires:

$$E_{bit} \geq k\,T\,\ln(2)$$

By compressing redundancies into invariants once, rather than repeatedly during inference, invariant-first AI dramatically lowers cumulative erasure cost and energy consumption.

6. Stability and Safety Considerations

Touching the compressed structural layer poses risks if constraints are treated as directives rather than biases. To prevent destabilization, invariant-first AI must adhere to the following safeguards:

- The invariant layer cannot generate outputs.
- Updates to invariants occur slowly and require multi-domain validation.
- Compression rate limits prevent rapid constraint shifts, analogous to protective pain signals in biological systems.

These measures align with UCST's emphasis on coherence preservation and prevent runaway self-modification.

7. Implications and Applications

7.1 Resource Efficiency

By prioritizing invariants, data requirements are reduced by orders of magnitude, model sizes shrink, and inference becomes cheaper and more stable.

7.2 Improved Prediction and Coherence

Invariant-first AI excels at long-range prediction, early detection of instability, and contextual understanding across domains, making it suitable for governance modeling, climate systems, and socio-economic forecasting.

7.3 Alignment and Ethics

Because invariants encode structural realities rather than values, ethical behavior emerges indirectly through coherence preservation rather than explicit moral rules, reducing alignment brittleness.

8. Conclusion

This paper argues that the escalating cost and instability of modern AI systems stem from an overreliance on instance accumulation and a neglect of compression-first design. Drawing on UCST and internal project research, we propose invariant-first AI as a viable alternative. By explicitly learning and preserving structural constraints, AI systems can achieve greater coherence, predictive power, and efficiency while reducing computational and energetic demands. Invariant-first AI represents a shift from remembering everything to remembering what cannot be violated.

References

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127–138.

Landauer, R. (1961). Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development, 5*(3), 183–191.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica, 14*(5), 465–471.

Shannon, C. E. (1948). A mathematical

theory of communication. *Bell System Technical Journal, 27*, 379–423.

Unified Consciousness Substrate Theory Research Group. (2023–2026). *Memory Bank and associated project files*. Unpublished internal manuscripts.