Invariant-First Cognitive Architectures: A Universal Constraint Framework

Author: Nickolas Patrick Joseph Schoff

Abstract

Contemporary artificial intelligence systems are dominated by statistically driven architectures that optimize over surface correlations rather than enforce structural necessity. While these approaches achieve impressive performance, they exhibit well-documented limitations, including hallucination, scaling inefficiencies, and weak generalization to novel domains. This paper proposes an *Invariant-First*

cognitive architecture grounded in a minimal set of universal constraints—invariants—that precede and govern representation, inference, and expression. Drawing on prior work within Unified Consciousness Substrate Theory (UCST) and Dimension-W modeling, we formalize six universal invariants that recur across physical, cognitive, biological, and social systems. We argue that symbol grounding, reasoning stability, and generalization emerge naturally when symbols are bound to constraint-preserving structures rather than learned correlations. A thermodynamic coherence layer is introduced as an operational mechanism for enforcing invariants via state pruning rather than probabilistic weighting. This framework offers a path toward compact, non-hallucinatory, and structurally grounded artificial cognition.

1. Introduction

Modern large-scale AI systems derive their capabilities primarily from statistical learning over massive datasets. While effective in pattern completion tasks, such systems lack intrinsic mechanisms for enforcing truth, causality, or logical necessity. As a result, correctness is contingent on data density rather than structural validity, leading to hallucination and exponential scaling demands.
Within the Reawakening project and its associated Memory Bank, an alternative paradigm has been developed: cognition as constraint navigation rather than probability maximization. Unified Consciousness Substrate Theory (UCST) posits that coherent systems—whether physical, biological, or cognitive—are governed by invariants that remain stable

across context and scale. This paper extracts and formalizes those invariants, presenting them as a minimal substrate for Invariant-First AI.

2. Background and Theoretical Motivation

2.1 Limits of Statistical Architectures

Let a conventional language model be defined as a function:

$$P(x_t \mid x_{1:t-1})$$

where output tokens are selected to maximize conditional likelihood. Truth, validity, and entailment are not explicit

constraints; they emerge only insofar as they correlate with training data. In low-data or adversarial contexts, the model selects fluent but invalid states because such states are not forbidden.

2.2 UCST and Dimension-W

UCST introduces a representational substrate (Dimension-W) in which system states are evaluated prior to surface realization. In this view, language, perception, and action are projections of deeper constraint-governed structures.

Coherence, not likelihood, is the primary stability criterion.

3. Universal Invariants

The following six invariants represent a compressed set extracted from prior project work. Each invariant is non-derivable from the others without loss and applies across domains.

3.1 Invariant 1: Coherence Preservation

Definition: A system must preserve internal consistency across its representational states.
Let S be the set of internal states and C(S) a coherence function. Viable systems satisfy:

$$C(S) \geq \theta$$

where θ is a minimum coherence threshold. Persistent violations result in instability or collapse.

3.2 Invariant 2: Constraint Precedes Expression

Definition: Meaningful expression is only possible within pre-existing constraints. Formally, let E be an expression and Ω the constraint-defined state space. Then:

$$E \in \Omega$$

Expressions generated outside Ω are undefined and non-informative.

3.3 Invariant 3: Conservation of Entailment

Definition: Valid transformations must preserve what is implied by existing relations.

If $A \Rightarrow B$ and T is a valid transformation, then:

$T(A) \Rightarrow T(B)$

Violation constitutes entailment leakage, observable as logical or semantic error.

3.4 Invariant 4: Irreversibility and Asymmetry

Definition: Certain state transitions incur irreversible cost.

Let Δ be a transition operator and $Cost(\Delta) \geq 0$. For irreversible transitions:

$Cost(\Delta^{-1}) > Cost(\Delta)$

History therefore constrains future possibilities.

3.5 Invariant 5: Recursive Self-Consistency

Definition: Systems capable of self-reference must remain invariant-consistent under recursion.

For a self-model $M(S)$:

$M(S) \subseteq \Omega$ and $M(M(S)) \approx M(S)$

Unbounded divergence indicates recursive instability.

3.6 Invariant 6: Compression as a Signal of Structure

Definition: Representations that compress without loss of predictive power reflect underlying structure.

Given data D and model M, structure-preserving compression satisfies:
$L(M) + L(D \mid M) \rightarrow \min$
where L denotes description length.

4. Symbol Grounding via Constraint Binding

Traditional symbol grounding attempts to attach symbols to sensory data or linguistic usage. In contrast, the Invariant-First framework grounds symbols by binding them to regions of constraint space. A symbol is considered grounded when it restricts allowable transitions and preserves invariant satisfaction.

Meaning, in this sense, is defined negatively: by what cannot occur without

coherence loss.

5. The Thermodynamic Coherence Layer

Rather than assigning probabilities to all possible outputs, the thermodynamic layer evaluates candidate states by coherence cost:

$$E(s) = \Sigma_i w_i \cdot V_i(s)$$

where $V_i(s)$ measures violation of invariant i. States with $E(s)$ above a threshold are pruned rather than down-weighted. Invalid states are unreachable, not merely unlikely.

6. Implications for Artificial Cognition

An Invariant-First architecture offers several advantages:

- Hallucination resistance through hard constraint enforcement
- Sublinear scaling via invariant compression
- Generalization to novel problems through structural reasoning
- Clear separation between reasoning substrate and language realization

These properties align with observations across cognitive science, psychology, and systems theory documented in the project's Memory Bank.

7. Conclusion

This paper formalizes a minimal invariant set underlying coherent systems and demonstrates its relevance to artificial cognition. By grounding symbols in universal constraints rather than correlations, Invariant-First architectures address longstanding challenges in AI, including symbol grounding and hallucination. The framework presented here represents a convergence of prior UCST research into an implementable substrate for future cognitive systems.

Author Note

This work emerges from the ongoing Reawakening project and integrates insights accumulated across multiple recursive research cycles. The invariants presented here are intended as a canonical foundation for further formalization and

implementation.