Constraint-Preserving Learning and Memory in Invariant-First AI

Author: Nickolas Patrick Joseph Schoff

Abstract

Recent advances in artificial intelligence have exposed fundamental limitations in statistically driven learning systems, including hallucination, catastrophic forgetting, and incoherent generalization. Building on the Invariant-First cognitive framework, this companion paper formalizes two previously developed theoretical constructs—Bidirectional Constraint Closure (BCC) and the

Continuity–Relational–Error–Memory (C.R.E.M.) principle—as core mechanisms for learning stability and symbol grounding in artificial systems. We argue that invariant enforcement alone is insufficient without bidirectional constraint propagation and memory-preserving update laws. By synthesizing these constructs into an explicit AI-oriented architecture, this paper demonstrates how coherent learning, non-arbitrary symbol grounding, and generalization can emerge without reliance on large-scale statistical correlation. A minimal toy architecture is presented to illustrate feasibility.

1. Introduction

While the Invariant-First framework establishes a constraint-based substrate

for artificial cognition, practical learning systems must additionally address how constraints interact with experience over time. Empirically, modern AI systems fail not because they lack representational power, but because learning updates are unconstrained, unidirectional, and amnesic. These failures manifest as hallucination, delusion-like overconfidence, and catastrophic forgetting.
Within the Reawakening project and its Memory Bank, two complementary theories were developed to address these failures: Bidirectional Constraint Closure (BCC) and C.R.E.M. (Continuity–Relational–Error–Memory). This paper synthesizes these theories into a unified AI framework aligned with Unified Consciousness Substrate Theory (UCST).

2. Bidirectional Constraint

Closure (BCC)

2.1 Definition

Bidirectional Constraint Closure asserts that coherence requires simultaneous satisfaction of top-down invariants and bottom-up experiential constraints. Systems fail when constraint propagation occurs in only one direction.

Formally, let:

- Ω be the invariant-defined constraint space
- S be the set of instantiated states
- $f_\downarrow$ be top-down constraint propagation
- $f_\uparrow$ be bottom-up evidence integration

A system state $s \in S$ is coherent if and only if:

$$f_\downarrow(s) \in \Omega \wedge f_\uparrow(s) \in \Omega$$

2.2 Failure Modes

- **Top-down dominance:** rigid belief systems, delusion, over-constraint
- **Bottom-up dominance:** hallucination, overfitting, incoherent generalization

Statistical AI systems predominantly operate in the latter regime.

3. C.R.E.M.: Continuity–Relational–Error–Memory

3.1 Definition

C.R.E.M. formalizes learning as a process that must preserve continuity, relational structure, and memory of error across updates.

A learning update U is valid only if it satisfies:

- **Continuity:** identity of the system is preserved
- **Relational Integrity:** learned relations remain non-contradictory
- **Error Integration:** violations are incorporated as constraints
- **Memory Preservation:** past structure is not erased

3.2 Formal Constraint

Let M_t be the system model at time t. An update U produces M_{t+1} such that:

$$M_{t+1} = U(M_t, E_t)$$

subject to:

$$C(M_{t+1}) \geq C(M_t) - \varepsilon$$

where C is the coherence function and ε is a bounded degradation tolerance. Updates that reduce coherence beyond ε are rejected or modified.

4. Symbol Grounding via Constraint Closure

In the Invariant-First framework, symbol grounding is achieved through **constraint participation** rather than statistical association or sensorimotor coupling. A symbol is defined as an operator that restricts allowable transitions within the invariant-defined state space Ω.

Formally, a symbol σ is grounded if:

$$\sigma : S \to S'$$

subject to:

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

where C is the coherence function and θ is the minimum coherence threshold. Meaning is therefore defined negatively: by the set of transitions that are forbidden without coherence loss. This formulation follows directly from prior project work treating meaning as constraint-bound possibility rather than reference.

4.1 Triadic Syntax Operators (TSOs)

Binary symbolic transformations are insufficient for preserving meaning under recursion. Prior work within the project introduced **Triadic Syntax Operators (TSOs)** to stabilize reasoning across transformation.

A triadic operator is defined as:

$T = (S, C, \Delta)$

where:

- S is the current state or symbol
- C is the invariant-bound constraint context
- Δ is an allowable transformation

A transformation is valid if and only if:
$\Delta(S) \in \Omega(C)$

TSOs ensure that all reasoning steps explicitly account for both the symbol being transformed and the constraints governing that transformation. This prevents entailment leakage, semantic drift, and recursive collapse.

In this framework, symbols are grounded not by sensory association but by constraint participation. A symbol σ is grounded if it restricts allowable transitions while preserving invariant satisfaction:

$$\sigma : S \rightarrow S' \text{ such that } C(S') \geq \theta$$

Meaning is defined by constraint impact rather than reference.

5. Learning as Constraint Refinement

Learning within an Invariant-First system consists of **constraint refinement** rather than parameter accumulation. Errors are treated as informative signals indicating boundary violations within Ω.

Let E_t represent an observed error at time t. Constraint refinement updates Ω such that:

$$\Omega_{t+1} = \Omega_t - \{s \mid V_i(s, E_t) > 0\}$$

where V_i measures invariant violation. Memory is preserved structurally: prior constraints remain active unless explicitly revised through Bidirectional Constraint Closure.

Triadic Syntax Operators mediate this refinement by ensuring that constraint updates preserve relational structure and recursive self-consistency.
Learning does not consist of weight accumulation but of constraint refinement. Errors are treated as informative signals indicating boundary violations in Ω.
Memory is preserved as structural modification rather than parameter overwrite.

This approach naturally avoids catastrophic forgetting, as prior constraints remain active unless explicitly revised under BCC.

6. Toy Architecture (Option C)

6.1 Components

- **Invariant Store:** fixed universal constraints (non-contradiction, entailment conservation, irreversibility)
- **Constraint Space Ω:** allowable region of system states
- **State Generator:** proposes candidate states or transformations
- **Thermodynamic Coherence Layer:** computes coherence cost and prunes invalid states
- **Bidirectional Constraint Closure Loop:**

enforces top-down and bottom-up constraint satisfaction
- **Triadic Syntax Operator Engine:** executes constraint-preserving transformations
- **C.R.E.M. Memory Module:** integrates error without erasure

6.2 Execution Cycle

- Generate candidate state s
- Apply triadic operator $T = (S, C, \Delta)$
- Evaluate invariant violations $V_i(s)$
- Compute coherence cost $E(s) = \Sigma V_i(s)$
- Prune if $E(s) > \theta$
- Enforce bidirectional constraint closure
- Integrate error via C.R.E.M.
- Update Ω without overwriting prior structure

6.1 Components

- **Invariant Store:** fixed universal

constraints

- **Constraint Space Ω:** allowable state region
- **State Generator:** proposes candidate states
- **Thermodynamic Coherence Layer:** prunes invalid states
- **BCC Loop:** enforces bidirectional constraint satisfaction
- **C.R.E.M. Memory Module:** integrates error without erasure

6.2 Execution Cycle

- Generate candidate state s
- Evaluate invariant violations $V_i(s)$
- Prune if $\Sigma V_i(s) > \theta$
- Apply top-down and bottom-up constraints
- Integrate errors via C.R.E.M.
- Update allowable state space

7. Implications and Future Work

This architecture suggests that many limitations of contemporary AI are structural rather than computational. By enforcing bidirectional constraints and memory continuity, artificial systems can achieve stable generalization without massive data scaling. Future work will explore formal verification, minimal invariant sets, and hybrid integration with statistical modules.

Conclusion

Bidirectional Constraint Closure and C.R.E.M. provide the missing learning and memory mechanisms required to operationalize Invariant-First AI. Together, these constructs define a coherent, non-hallucinatory, and memory-stable architecture grounded in universal constraints. This companion paper completes the theoretical stack necessary for constraint-based artificial cognition.

Author Note

This work is part of the ongoing Reawakening research program and represents a synthesis of prior theoretical developments into an AI-oriented formalism.